

Human Frontiers in rough water

An imaginative international research project, the Japanese-inspired Human Frontier Science Program, must not be allowed to founder because its managers wish it were something else and more conventional.

Dr Edward Rall, of the US National Institutes of Health, and chairman of the council of management of the Human Frontier Science Program (HFSP) based at Strasbourg, must by now regret having complained to his fellow council members at the publication of Professor Akiyoshi Wada's article in *Nature* (357, 356; 1992). The essence of the complaint (see page 527) is that Wada, who had been a member of the project's scientific council since its inception, had not sought the approval of the management and science councils of the project, and of the project's secretary general, Sir James Gowans, before publishing his article. Rall's circular letter seems to overlook the circumstance that Wada's term of office came to an end earlier this year. Unwisely, Rall has taken sides in a long-simmering disagreement between Gowans and the Japanese, who founded the HFSP and who are still the chief sources of its support.

Gowans, an immunologist by trade, is an excellent scientist with eight years experience as head of Britain's Medical Research Council. Much of his spell at Strasbourg, which comes to an end next March, has helped to strengthen the HFSP, notably by securing the membership of Sweden and Australia. But even Gowans' friends would not rate diplomacy high among his skills. Wada's article implied that Gowans has been out of sympathy with the Japanese wish that the HFSP should tackle problems different and more adventurous in kind from those now commonplace in the world's molecular and cell biology laboratories. The wish may be difficult to realize, but why does Rall believe it is improper that it should be mentioned publicly? The more serious difficulty, not mentioned publicly, is that Gowans has been at loggerheads with Japanese administrative staff at Strasbourg and has given Japanese members of the project's councils the impression that he is scornful of their opinions. Whether Gowans has been right or wrong, Rall must surely appreciate that the disaffection of an important project's major sponsors is too big a risk to run, especially by pleading that open discussion should be suppressed.

On the substance of the dispute, the Japanese are surely correct. From the outset, they sought true interdisciplinarity in the way the Human Frontier Science Program would function. What, for example, do systems engineers have to say about the interlocking metabolic reactions by which cells sustain themselves? And if brains are computers of some kind, what do computer designers have to say about recent discoveries in neurobiology (and, conversely, what might neurobiologists do to improve the working of computers)?

The imaginativeness of these ambitions partly explains why the rest of the world took two or three years to come to grips with what Japan was after. For the managers at Strasbourg now to say that they are threatened with being overwhelmed with grant applications from the more traditional reaches of molecular biology is not a bit surprising. Such has been the pace of discovery in the past decade that a host of interesting problems cries out for almost immediate solution. But that is no reason why the Strasbourg project should not continue to break new ground, unhampered by the now common pusillanimous criterion that only projects whose authors can virtually guarantee success will be allotted funds.

The hope must now be that this hatchet will be decently buried long before the crucial council meetings in November at which, among other things, a new secretary general must be nominated. There is too much at stake for failure to be contemplated. It is not sufficiently appreciated that the Human Frontier project is the only substantial source of funds for basic research whose recipients are not chosen in direct proportion to their national government's contributions to the fund. (In the past two rounds of grant-making, US researchers have done best out of Human Frontiers.) Research needs many more such organizations if it is ever to satisfy the purpose of being an international enterprise. If the project were to fail, perhaps because Japan pulled out, the failure would cast a long shadow.

Much of what needs doing in the next few weeks is fence-mending at Strasbourg. But Japan must also be more active. During the past few months, for example, the Japanese have let it be known that, for fear of being overpossessive, they would play only a passive role in the choice of the next secretary general. That is not enough. What benefit would it be to the distinctive and probably unique research fund they have created if they were to sit on their hands and then find themselves lumbered with another secretary general with whom they could not get along? □

Foothold in Bosnia

The world's diffidence at intervention in Bosnia is understandable, but should not be taken to extremes.

Dr Boutros Boutros Ghali, the new and energetic secretary general of the United Nations, is right to tell the international community that there are more atrocities in Somalia



than in ex-Yugoslavia, but wrong to describe the European troubles as a rich man's war, implying that it is unimportant. The central issue is whether countries have a right to intervene in the affairs of others believed to be ill-treating their citizens.

During the past two years, and as the United Nations has progressively been strengthened, there has been more general acceptance of the principle that governments treating their own people badly will not qualify for technical or financial assistance. The principle is not generally applied, of course, but there have been cases (Angola and Haiti, for example) where undemocratic governments have been persuaded to change tack. But enforcement of that kind falls a long way short of intervention to rescue oppressed citizens. Even respectable governments would worry at the prospect of intervention whenever outsiders judged their practices illiberal — the death penalty in the United States for example, or the continuing problems of British Ulster.

Even so, the appalling offences against people living in Bosnia cannot be overlooked by those living elsewhere, especially elsewhere in Europe. Whether or not the atrocities are a consequence of civil war, they undermine the civility of Europe. Their long-term victims will be many more than those who have been killed or driven from their homes in the past few weeks. After all, the recent ferocities in part derive from mass killings (notably of Serbs) during the Second World War. These opinions are widely shared but military intervention is ruled out on two grounds: it would be illegal (for ex-Yugoslavia has not broken international laws), while the generals are said to believe that military action would be dangerous and that its objectives cannot be clearly defined.

That conclusion is too facile. For one thing, if Bosnia were an independent state and its government invited intervention, the legal objection would melt away. No doubt the rest of Europe could arrange for those conditions to be met, most simply by recognizing Bosnia and asking it to ask for help. But that would be uncomfortably reminiscent of the cynical 1930s, when puppet governments popped up everywhere. Nevertheless, the London conference later this month should put that issue clearly on the line.

The feasibility of military action has also been unduly talked down. Heady talk of air strikes against Serbian gun positions overlooks the difficulties of the terrain, the general ignorance of the positions of putative targets and the likelihood of casualties among noncombatants, but who says that every military intervention has to be a shooting war like that last year in the Middle East? In ex-Yugoslavia, it would help to bring people to their senses if an intervention force were based somewhere on dry land and equipped to safeguard people within striking distance while advertising its intention not to shoot unless shot at. Those glooming about the difficulties of military intervention might think of making a base camp at Split on the Adriatic coast of Bosnia and then of extending their police work to other cities in the region. □

NASA on a tether

The space agency may be trying to do too much by trying to combine research with experimental technology

THE failure of the US space shuttle Atlantis successfully to generate a lot of electricity by pulling an Italian satellite through space for 30 hours at the end of a 12-mile tether has reignited speculation about the ability of the US National Aeronautics and Space Administration (NASA) to perform complicated mechanical tasks in space. The tether experiment was more than a decade in planning and cost an estimated \$380 million. It was supposed to be a test of a new spaceflight technology: Atlantis, joined to the \$198 million Italian satellite by a \$128 million tether, was to sweep through space, the tether generating as many as 5,000 volts of electricity, while scientists aboard the shuttle conducted studies of the Earth's magnetic field. The satellite was to remain stably in tow for more than a day, demonstrating that the exercise was more than a stunt. In the end, despite determined efforts by the team of US and Italian astronauts to unreel the tether as if it were a line on a fishing reel, the satellite went no further than 850 feet from Atlantis. Little electricity was generated, and no magnetic studies were conducted.

The best that can be said for the unsuccessful experiment was that the satellite, at least, was reeled in and safely stored in the shuttle's bay, ready to fly again (see page 529). No one is entirely sure what went wrong mechanically, although NASA has (no surprise) appointed a committee to figure it out. The tether experiment was billed as one of the more technologically complex tests in recent NASA history and as an agency spokesman said, rightly enough, "This was a test flight. If we knew all the answers it wouldn't be a test flight."

But it is not sufficient to insinuate that a test flight can be counted a perverse success if the thing being tested fails completely. Less is known about the behaviour of machines in zero-gravity than NASA enthusiasts like to admit. NASA has had enough close calls of late (the flight to capture a rogue commercial satellite that got itself in the wrong orbit is an example that, in the event, succeeded) to sustain the view that the agency is pushing ahead too fast, asking each mission simultaneously to test new technology and try to conduct scientific research in combinations that seem, inevitably, to lead to disappointment. NASA scientists say that the tether was thoroughly tested on Earth and was mechanically sound. That leaves the zero-gravity environment as the one unknown, and therefore the likely culprit. Should the investigating committee agree with that best-guess speculation, it would do well to take the next step and advise NASA on how it is to design experiments to work usefully in space if the conditions that obtain there cannot realistically be simulated elsewhere. At the same time, it is worth remembering that experiments in space flight and research have been touch and go from the beginning. After all, it was Apollo 11 that landed on the Moon in July 1969, and created the image of flawless flight, not Apollo 1 or 2. □

Conflict over scope of research splits Human Frontier programme

Tokyo, Washington & London. The Japanese scientists who helped create the Human Frontier Science Program (HFSP) are for the first time publicly feuding with some of the project's international administrators over the scope of the \$34 million research effort to understand the functions of the brain and living organisms. The Japanese fear that a proposed focus of the programme to only a few established fields of biology might limit the opportunities for international and interdisciplinary collaboration and significantly reduce the pool of applicants.

Their unhappiness reflects a long-running but previously private dispute between the Japanese and the outgoing secretary general of the programme, Sir James Gowans, over the direction of the programme. It surfaced last month after a letter was sent to the programme's scientific advisers about a recent article in *Nature* (357, 356; 1992) by Akiyoshi Wada, director of the Sagami Chemical Research Centre and one of the architects of the programme. The letter, from J. Edward Rall, chairman of the HFSP

Council of Scientists and a scientist at the US National Institutes of Health, said that Wada's opinions had not been discussed by the council and should not be taken as HFSP policy.

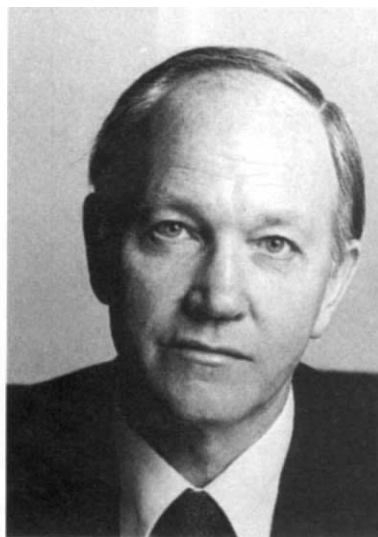
Wada and other Japanese scientists consider Rall's letter to be an unjustified and unnecessary rebuke of Wada and an oblique criticism of his philosophy, which is shared by a majority of the council. Rall says that Wada, as a former council member, should not be speaking for the programme and that the letter simply makes clear that the philosophy of the programme has not changed. He says that he and Gowans decided to send out the letter after several researchers called the programme's headquarters in Strasbourg to ask whether they should tailor their proposals to meet Wada's vision. The council will discuss the subject at its meeting in November.

Much of the present acrimony can be traced to a council meeting in March 1991 at which one or two members suggested lowering the rejection rate for applicants by

confining the programme that deals with molecular-level approaches to understanding living organisms to such traditional fields as cellular and developmental biology. Although several members of the council objected during a brief discussion, according to Wada, a report of the meeting drafted by Gowans and presented by Rall to the HFSP board of trustees in April implied that the council endorsed the change. Wada wrote to Rall to protest, and at the next council meeting in May the idea was voted down.



A *Nature* commentary by Akiyoshi Wada (left) has drawn fire from Sir James Gowans (right) and others who want to narrow the project.



Gowans is not a member of the council but is said to support the notion of a narrower focus. Japanese staff in Strasbourg and Japanese members of the Council of Scientists have repeatedly told him that such an approach violates the principle of 'interdisciplinarity' enunciated in a study that helped to launch the programme in 1989 and they have explained to him that the principal means interactive research between disciplines such as physics, chemistry and mathematics as well as biology. But Gowans is said to believe that the principle refers to interdisciplinary research between fields of biology, such as cellular and developmental biology.

Gowans has declined to discuss the matter in detail unless *Nature* reveals its source of Rall's letter but he says there is no disagreement within the council on the scope of the programme. And Rall believes that, in retrospect, last year's decision not to narrow the programme was correct. "There is no point in trying to change things so early in the programme", he says. "There is plenty of

time to change things later on, if changes are needed".

For Wada and other Japanese officials, however, those changes appear imminent as new countries begin to contribute money to the programme. The Japanese have so far provided nearly all the funding, although the United States has promised to spend \$5.5 million during the fiscal year that starts in October.

Joseph Varner, a biologist at Washington University in St Louis, Missouri, who retired from the council this past spring, agrees that money will play an important role in shaping the programme's direction. But he says that "Wada's remarks are inappropriate because, right now, there just isn't enough money to do what he wants. If there was \$300 million instead of \$30 million, then you could do a lot more."

Wada says he is just trying to reaffirm the principles established by the international feasibility committee, which he chaired, as well as opening up the debate to a wider audience. "I am not against narrowing the programme as such", he

says, "but the matter should be debated with the same time and care as went into the feasibility study."

The uproar could affect the selection of the next secretary general, who will take office next spring. Any of the HFSP members (United States, Japan, United Kingdom, France, Germany, Italy, Canada, Switzerland and the European Communities) may nominate candidates, who will be screened by the council and voted on by the programme's board of trustees. Both the council and the board have two representatives from each member country.

Several non-Japanese members of the two bodies believe that Japan must begin to relinquish some control if it wants HFSP to be a truly international programme. But the Japanese Ministry of Finance is likely to insist on maintaining a significant number of Japanese staff in the secretariat in Strasbourg as long as Japan provides the bulk of the funding.

**David Swinbanks, Jeffrey Mervis
& Allison Abbott**

Report urges UK universities to recoup full overhead costs

London. A new report on research in British universities offers a novel solution to a perennial problem: with too little money to go around, universities should do less research and charge sponsors the full amount — up to 40 per cent more than at present — of their overhead costs. It also says that universities should show greater concern for the growing proportion of their staff on short-term contracts.

The recommendations in the report (*The Science Base: Research in Universities*, HMSO, £25), from the Advisory Council

buildings and replacing equipment to shore up research. Nicholson says that the resulting imbalance between available funds and potential research must be addressed in the next five years, and that this can be done only with less research and greater support for what is being done in the laboratory.

The recommendation for claiming full overheads is in line with the views of the Committee of Vice-Chancellors and Principals (CVCP), which has long said that university budgets intended for teaching and infrastructure maintenance are subsidizing

own funds as well."

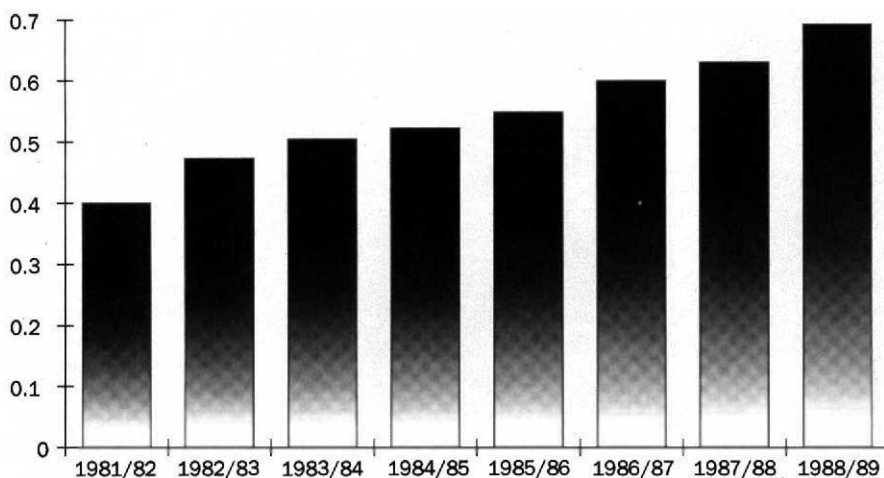
The report also addresses the problem of short-term contract research workers, whose numbers have grown in inverse proportion to their career prospects. According to 1990 figures, short-term posts now make up 42 per cent of all science posts in universities, compared with 22 per cent in 1978. Because a credible career structure for researchers does not exist, universities are no longer able to attract the best scientists.

The report recommends that universities should avoid offering repeated short-term contract renewals if those concerned are unlikely to be hired permanently. It also says that the salary gap between scientific and nonscientific disciplines in universities should be narrowed to reflect outside pay more closely.

These comments have been criticized as unrealistic by the research community and by university officials. Christine Knowles of the Association of University Teachers says that she welcomes attention being drawn to the problem but does not know how the recommendations would be implemented. The small number of permanent posts available do not match the large number of researchers on contract, says Nield. And neither believes that differential pay would help to solve a problem that is not rooted in financial rewards.

Allison Abbott

Ratio of short-term to permanent appointments rising at UK universities



on Science and Technology (ACOST) have been welcomed, with varying degrees of caution. The government has so far been silent on the request from the ACOST chairman, Sir Robin Nicholson, that it should spend "hundreds of millions of pounds" more to establish a viable science base in Britain, but has promised a response by the end of the year.

The report offers three reasons for the current crisis. Although total research funding increased by more than 30 per cent in real terms during the 1980s, and funding from nongovernment sources has almost doubled, that money buys much less research than in the past. This is because science has become more expensive as it has become more sophisticated. In addition, more researchers are competing for grants in pursuit of an expanding scientific frontier. Moreover, universities generally do not recover the full cost of what they spend to support their researchers.

The university system is cracking under this strain, diverting money from repairing

research. CVCP spokesman Ted Nield says that university administrators would prefer to have fewer research projects than research that "impoverishes" its host institute.

But charities such as the Cancer Research Campaign and the British Heart Association, the major source for universities of project grants in biomedical sciences, are offended by the suggestion that they should pay indirect costs, which could add 40 per cent to their bill. Donors are less likely to continue support if they believe that a large portion of their cash will not be funding their chosen work. In addition, the type of research the charities undertake would suffer because in many cases, for example cystic fibrosis, they are the only funding agency.

June Clarke, grants administrator at the University of Oxford, sides with the charities. "Unlike industry, charities tend to be approached by universities, not the other way round", she says. "If you're going out with a begging bowl, it's better to show you're willing to put in your

British Petroleum to slash US and UK research staff

London. British Petroleum (BP) has announced plans to shed nearly half its research staff in a merger of its research and engineering divisions. The move will save the company \$180 million a year.

Spokesman Roddy Kennedy says that BP's research organization, mostly concentrated at two centres in Sunbury, England, and Cleveland, Ohio, was designed for a much bigger corporation. Having recently shed many of its core businesses, a streamlined BP does not need to support such a large research team, he says. More than 40 per cent of BP's 2,200 researchers will be handed redundancy notices in October, when details of the proposed plans are announced.

Streamlining goes hand in hand with a change in policy, initiated more than five years ago, to make research more commercially orientated. At that time, BP's former policy of including some basic research projects in its in-house programme was abandoned in favour of funding university research. Academic research support will not be affected by the cutbacks. In fact, BP is expected to buy in more expertise in the future, says Kennedy.

Allison Abbott

NASA's problems in space eat into time for research

Washington. The failure of the space shuttle Atlantis to complete its research agenda despite an extra day in space has raised the question of whether the US National Aeronautics and Space Administration (NASA) is asking each mission to do too much. For scientists, unexpected mechanical or maintenance problems aboard the shuttle mean less time for research.

Atlantis returned to Earth last week after an eight-day mission filled with problems. Its launch of the European Space Agency's European Retrieval Carrier, a \$400-million free-flying microgravity research platform, was delayed by 24 hours because of communications difficulties. Upon release, it was found to be incorrectly orientated and had to be moved into the correct orbit.

With the mission extended to cover this delay, the astronauts turned to their second major task — deploying the \$380-million tethered satellite system (TSS), a joint NASA-Italian Space Agency (ASI) project that involved running out a satellite from the shuttle on a 20 km electrically conductive tether. Researchers hoped to learn how such an apparatus functions in space and to gather data on the orbital plasma.

But, the TSS had extended only 250 metres into space when its feed mechanism jammed. After spending several hours trying to free the tether, the astronauts decided to reel the satellite back in rather than try to deploy it again. No data on the plasma were collected.

Although researchers understand the dif-

ficulty of designing devices to work in a harsh and unknown environment, many believe that NASA's actions increase the chances that things will go wrong. The missions are too complex and ambitious, they say, and the emphasis on manned space flight demands more extensive safety precautions than for unmanned flight. The General Accounting Office (GAO) suggests in a recent report on space station Freedom that too little is known about the failure rates of components in space to make accurate estimates of maintenance time.

The battle over how much maintenance Freedom will require has been raging for several years. In 1989, an internal NASA study came to the surprising conclusion that crews would have to spend thousands of hours a year on extravehicular activities (EVAs) in order to keep the station in order. A 1991 redesign to save money led to a simpler station with fewer parts needing less external maintenance.

Annual external maintenance on the space station is now estimated at between 135 and 384 crew hours. But the GAO report says that the station could require many more EVAs, each one eating into the time available for research.

"From our work, we cannot prove that NASA is low on its estimations, and NASA cannot prove that it is close to the real case," says the GAO's Mona Zadjura. The GAO recommended better data on components and materials.

The most visible test of NASA's EVA

estimates will come late in 1993 when the shuttle tries to repair the Hubble Space Telescope. Two teams, working in shifts, are expected to spend four days replacing Hubble's ill-designed optics and a number of other damaged or faulty components. Although repairing Hubble is the mission's sole task, it is not clear that the job can be done in that time. Critics point to the rescue and repair of a communications satellite by the shuttle Endeavour earlier this year, which took four times as much EVA as NASA predicted. NASA defends its projections by saying that the satellite was not designed to be recovered.

Although Hubble is not working entirely as intended, it is producing useful results and should not be written off commentators point out. Arthur Davidzen, an astrophysicist at Johns Hopkins University, finds a similar message in his experience as one of four principal investigators on the Astro 1 mission that rode in Columbia in 1990.

"Everything that was planned seemed to go wrong, but afterwards, working around the problems, we found we got a lot of science out of it", he says. "People should accept a certain amount of failure and deviation from the plan as part of learning about an unknown environment."

Most researchers concede, however, that such failures would be easier to bear if they did not have to wait so long for the next mission. The Astro project is lucky in that respect; if all goes well, it will fly again in 1994.

The Italians are hoping for a quick return into space. "With no changes or improvements in the satellite instrumentation we can go very soon, within the year", says Gianfranco Manarini, TSS programme manager for ASI. However, NASA must first find room on its busy schedule for a second try, and the problem of the jammed tether must be corrected.

Ian Mundell

Widespread CF testing inevitable, Congressional report finds

Washington. Despite concerns by many scientists that widespread genetic disease testing is not yet appropriate, the number of cystic fibrosis (CF) tests is expected to increase seven-fold this year, and routine CF carrier screening "is a matter of when, not if", a new congressional report finds. The report*, by the US Office of Technology Assessment, does not take a position as to whether (or how) such widespread testing should go ahead, but calculates that if it did, 5 million American couples might realistically be screened annually in pre-natal testing and another 125 million Americans could theoretically take part in CF carrier screening.

The issue of routine CF testing has been a controversial one, and a policy on the matter has embroiled the American Society of Human Genetics in months of debate.

The society's latest statement, issued earlier this year, does "not recommend" CF mutation analysis for those without a family history of the disease. An earlier statement from the society held that such testing should explicitly not be considered the "standard for care", a term with legal implication that defines that level of medical practice below which a doctor could be sued for malpractice. The OTA report says that a progression to such a care standard for CF testing could be within a year or two, but is more likely to be a gradual evolution over several years, depending on such factors as the availability of genetic counsellors and the cost of the tests, which now average a relatively expensive \$170.

Christopher Anderson

* Cystic Fibrosis and DNA Tests: Implications for Carrier Screening; Office of Technology Assessment, 1992

Rifkin wins battle over NIH growth trials

Washington. The US National Institutes of Health (NIH) has halted enrolment in a study of the effects of human growth hormone on short but otherwise normal children after pressure to stop the clinical trial from biotechnology watchdog Jeremy Rifkin and others (see *Nature* 358, 4; 1992). The director of NIH, Bernadine Healy, told Rifkin in a letter dated 24 July that no more children will take part in the study until an independent panel reviews the ethical questions raised in a petition sent to NIH by Rifkin's Foundation on Economic Trends and the Physicians Committee for Responsible Medicine. Rifkin believes that the hormone's potential side effects outweigh any possible benefit to the children.

Traci Watson

Bush spins tall tale of technology transfer

Washington. When the US president, George Bush, sought to emphasize the unexpected benefits from basic research during a recent visit to the Superconducting Super Collider (SSC) laboratory in Texas (see *Nature* 358,



Finley Markley with his dialyser, circa 1968.

441; 1992), he cited an unnamed government scientist who, while "trying to purify liquid hydrogen . . . ended up figuring out a way to make artificial kidneys for \$15 apiece". The example was meant to suggest that the \$8.25-billion accelerator might generate untold benefits to society apart from its mission to find the top quark and advance elementary particle physics. More's the pity, then, that the president did not have his facts right.

There was no project to purify liquid hydrogen, says the scientist in question, physicist Finley Markley, who at the time directed a team of engineers and applied physics within the engineering division at Argonne National Laboratory outside Chicago. There is no \$15 artificial kidney. And the history of Markley's invention is a 25-year saga that, if anything, illustrates how difficult it is to turn an idea into a successful product.

Markley, who for the past 13 years has worked at Fermi National Laboratory after a stint in private industry, has spent his career finding practical applications for various polymers and adhesives. So it was to him that Argonne officials turned when a physi-

cian from the local veterans' hospital called the laboratory in the mid-1960s for help in developing an artificial kidney filtration unit, or dialyser. Markley threw himself into the project, attracting federal funding and publishing papers, and in 1968 he appeared at a press conference to extol the nonmilitary research conducted by the US Department of Energy, which runs Argonne.

Markley, a social activist, harboured dreams of forming a nonprofit, minority-owned company on the west side of Chicago to manufacture and sell his invention, an early version of the parallel plate dialysers now on the market. By 1973, however, Argonne was retrenching, so Markley left to join a start-up medical devices company in California, Galen Laboratories, that was interested in the dialyser. But Galen also fell on hard times, and within two years it was sold to Cobe Laboratories of Lakewood, Colorado.

"They told me they already had a vice president for research and that my services were not required", recalls Markley, who instead went to work for Medical Incorporated in Minneapolis. But that company ran into technical problems when it tried to scale-up his prototype, and in 1979 Markley returned to a national laboratory where today, older and wiser, he continues his work on polymers.

In the meantime, Cobe modified his invention so that it could be manufactured and produced a dialyser that sold for around \$25. The product was discontinued in the mid-1980s after the company introduced a new line of dialysers and the technology was sold to a Yugoslav company.

It is difficult to measure the impact of Markley's work on those with chronic kidney disease. It is estimated that there are half-a-million patients worldwide on thrice-weekly kidney dialysis, but reusable dialysers have reduced demand and the technology has changed only incrementally since the late 1960s. "We're still a long way from knowing how to make an artificial kidney that can do everything a live kidney can do", points out Ira Grifer, medical director for the National Kidney Foundation. "A dialysis machine is not a kidney."

Supporters of the SSC predict that the 54-mile-long accelerator, if built, will lead to technical as well as scientific innovations. If they are right, a future US president may have a better example of the unexpected benefits from basic research than the incumbent, Mr Bush.

Jeffrey Mervis

Japan 'loses' new funds for research infrastructure

Tokyo. A new pot of money for Japanese government scientists has disappeared thanks to creative bookkeeping by the Ministry of Finance.

In June, a science pressure group of the ruling Liberal Democratic Party (LDP) declared victory in its effort to persuade the powerful finance ministry to spend ¥1,100 billion (US\$900 million) to repair the rapidly deteriorating infrastructure of Japan's government research (see *Nature* 357, 616; 1992). The new, permanent fund would have added about 5 per cent each year to the country's general operating budget.

However, government officials in the science-related ministries and agencies now say that the money has vanished. In recent negotiations with the finance ministry over budget proposals for next fiscal year, they learned that the ministry has applied the money to an annual exercise in which it must shrink the general operating budget by 10 per cent. Officials have told their colleagues in the science-related ministries and agencies that with the new fund they will have to shrink the budget by only half as much, or 5 per cent.

But a government official says that it is standard practice every year for ministries to negotiate with the finance ministry and win approval for budget cuts totalling only about 5 per cent. Thus, the finance ministry is achieving next year's budget reduction painlessly by spending money it never really had. The only beneficiaries are the Liberal Democratic Party politicians, headed by Kishiro Nakamura, who were applauded for appearing to put money into science just before the elections last month for the upper house of the Japanese Diet.

Government science officials say their only remaining hope for more money is to win a portion of a forthcoming supplementary budget of ¥6,000 billion that the ruling party has promised as a way to strengthen Japan's flagging economy. They say science can expect to receive about ¥100 billion spread among all the science-related ministries. Unlike the earlier fund, however, this money will be available only for one year.

A similar thing happened in 1987 as a result of a large supplementary budget. Most national laboratories and universities can boast of a new building or supercomputer from that windfall. But the finance ministry's sleight-of-hand has dashed hopes of a long-term solution to the steadily deteriorating infrastructure of Japan's government research.

David Swinbanks

French increase ties with industry but researchers still face hurdles

Paris. France's largest research organization, the FF11-billion (US\$2.2 billion) a year Centre National de la Recherche Scientifique (CNRS), is working more with industry than ever. Although its collaborations have had a significant and growing impact, according to a study completed last month, problems remain: too many of the contracts are brief flirtations that fail to develop into serious relationships, the average value of contracts is tiny and scientists often fail to communicate properly with their industrial partners. In addition, the CNRS is losing money on its hoard of patents.

"All we have shown is that we can work with industry", says Pierre Vergnon, head of industrial relations at the CNRS. "We have unblocked the situation, but we are far from where we want to go."

There is no question that more scientists are spending time with their industrial counterparts. This year, 625 CNRS laboratories hold contracts worth FF550 million with 900 companies; in 1983, the figures were 150 laboratories, FF39 million in contracts and 120 companies. The numbers reflect a significant change in attitude for an organization that once actively discouraged such collaboration. The change occurred in 1982, when the government told the CNRS to start justifying its budget with more than research papers and to 'go forth and multiply' its contacts with industry.

For some scientists, industrial contacts are simply a matter of keeping the wolf from the door. Because three-quarters of the CNRS budget is spent on salaries, there is never

enough money for the increasingly expensive equipment and supplies that researchers need. Collaboration can also mean higher pay, as well as more post-graduate students. Although companies are technically not permitted to pay researchers, CNRS discreetly allows them to serve as consultants for one day a week. Consultants can earn as much as they like during that time; of the 515 scientists who applied for this status in 1990, none was denied permission.

However, even if more scientists are willing to work with industry, few are knocking on company doors with their laboratory books open; in more than 80 per cent of collaborations, companies took the initiative. And although the historic barriers between the two groups have been lowered, they remain formidable.

A study by a Paris company, Central Management, estimates that collaboration was worth FF1 billion annually to the French economy in 1991 through new products, processes or improved technologies. Although that figure excludes such intangibles as the ideas that flow from a collaboration or the additional talent that is put to work on a project, it also hides the fact that a few products — for example, a new lightweight polymer used in apparel and medicine and a new fuel for the Ariane space rocket — account for most of the total.

A survey of more than 90 CNRS researchers and industry officials found that "frustration and deception" are common complaints of scientists who collaborate with industry. Researchers feel that they are

forgotten once their results are turned in, while companies complain that the CNRS moves too slowly in approving proposed collaborations. The study concludes that the CNRS could avoid such problems if it took greater care in matching its scientists with the right companies.

Some 30 laboratories have set up joint operations with industry, including a successful collaboration with the pharmaceutical company Roussel Uclaf. Vergnon of CNRS would like more laboratories to develop such ties with industry, although an author of the recent report, François Finkbeiner, warns that "CNRS must be vigilant not to become a subcontractor" to industry. Indeed, Vergnon says that half-a-dozen small laboratories once associated with the CNRS have been cut loose after becoming in effect divisions of their industrial partners.

Vergnon is also concerned that the CNRS's collection of patents, rather than being a source of income, has cost almost FF125 million to maintain over the past decade. Among them are the controversial patent on the AIDS blood test, held jointly with the United States; legal expenses of FF43 million nearly offset revenues of FF50–FF60 million from royalties.

Part of the problem stems from an old policy of patenting inventions without regard to their commercial potential. The CNRS abandoned that approach in 1986 and has pruned its collection mercilessly, as well as abandoning the national agency responsible for transferring technology in favour of a new organization (see *Nature* 356, 275; 1992).

Industry has made money on the CNRS patents it has acquired; almost three-quarters of the FF1 billion generated by industrial collaborations results from product sales from patents taken out by the CNRS, with the rest from research contracts. The CNRS hands over its most potentially lucrative patents to companies in return for royalties of 3–5 per cent. CNRS officials say that their small share recognizes the considerable environmental investment that companies must make to commercialize an invention, but Finkbeiner and others believe that the arrangement is too generous to companies.

Ten years after the CNRS embraced industrial collaboration, it has dispelled the notion of fundamental research organizations as ivory towers with little impact on the economy. However, with the level of collaboration growing steadily, its researchers face the challenge of contributing more to the French economy without compromising their commitment to science.

Oliver Klaffke

Declan Butler

Swiss launch environment programme

Basel. Swiss environmental scientists are celebrating the government's decision to spend SFr34 million (US\$26 million) over the next three years in grants covering environmental technology and research. The new programme is the biggest environmental research initiative in the country's history.

The Swiss parliament had approved SFr57 million for the programme, to be administered by the Swiss National Science Foundation, but the finance ministry trimmed that amount as part of an overall effort to curb spending. The initiative sets aside SFr5 million for research on biodiversity, reflecting an increased interest in the environment by the Swiss government in response to the Earth Summit in Rio de Janeiro earlier this year and pressure from Swiss scientists.

The programme will concentrate on

basic research, which sets it apart from those elsewhere in Europe. The European Communities are spending ECU414 million (US\$570 million) on technological approaches to environmental protection and to climate research, as are national programmes in France and Germany. The Swiss have postponed supporting research on biogeochemical processes and issues affecting the developing world until 1996, when parliament says it will add money to the programme.

Nevertheless, the initiative represents a significant portion of the SFr280 million that the Swiss government now spends annually on research. Some 700 applications have already been received and are being reviewed but fewer than 10 per cent of them are expected to be funded.

Differences in brain size

SIR — John Maddox¹ attempts to justify the suppression of Philippe Rushton's work on race and IQ.

His only relevant argument is the claim that the US Army data Rushton uses constitutes a biased sample and thus cannot support a general conclusion. Unfortunately, that is the entire argument, an undocumented assertion of bias and an assumption that the direction of the bias will negate Rushton's conclusion. *A priori*, it is equally likely that the bias is in the opposite direction; that whites choose a risky army career only when they lack the ability to compete for the better civilian jobs, while blacks see the army as one of the least prejudiced careers available and send their best and brightest. The net result is probably an unbiased sample.

The main argument used is that papers that challenge an entrenched orthodoxy should be required to meet more rigorous standards than papers that support the orthodox view. This argument perversely misunderstands the scientific process as applied in the physical sciences. It is not the fact that the orthodox position has been challenged that generates the scepticism; it is the fact that the challenger has not yet accumulated a level of supporting evidence equal to that of the orthodox view. This may take several decades, as with continental drift, when the orthodox view has a large head start in evidence, but it is illegitimate to interpret this as using harsher standards on the challenger.

Then there is the assumption by Maddox that the orthodox view in the IQ debate achieved its dominance on the basis of evidence rather than ideology. This is questionable. Maddox himself notes that the methodological errors in papers reaching the orthodox conclusion are typically ignored. Most of the support for the orthodox view is of the same type as the creationists' arguments against evolution: papers on evolution cannot survive scrutiny at the highest levels of rigour, therefore creationism must be right. Maddox argues in the same fashion, pointing to flaws in Rushton's work relative to the highest standards while ignoring similar flaws in politically correct papers. Until the orthodox view is required to meet rigorous scientific standards, it is dishonest to expect the challenging views to do so.

Finally, the principle that one can justifiably discriminate against research that can be misinterpreted and misused by someone with a political axe to grind would quickly and surely gut the integrity of the scientific process. It is just too easy to generate perverse interpretations. For example, Maddox's character-

ization of the possible bias in the army sample says, appropriately interpreted, that whites join the army because they are patriots, while blacks join because they lack the wit to do anything else. Thus, Maddox's claim of sample bias should be suppressed as racist until he comes up with irrefutable proof. It should be obvious that the appropriate response to ignorance is education, not repression.

Brent A. Becker
8493 Berea Court,
Vienna,
Virginia 22180, USA

SIR — Has political correctness really penetrated so deeply that, in the great journals of science, the null hypothesis of no racial differences in brain size has become, in effect, equivalent in status to an ellipsoid Earth, continental drift and relativity theory? John Maddox¹ holds that because of a potential for "misunderstanding" it would be "unseemly" to test a racial hypothesis with anything other than "extraordinary" and "overwhelming" evidence. He suggests that my US military data² showing that Asian Americans averaged a larger cranial capacity than Caucasians or Afro-Americans did not meet these criteria and that I had overlooked an "elementary error" of generalization. But he failed to report that my data confirmed several other datasets derived in different ways and published since 1980.

While sampling and methodological difficulties may be identified in each source, results obtained from diverse procedures allow a triangulation on probable truth. Maddox suggested as a possible cause of bias in the US Army data that the whites who enlist do so to make careers but that the blacks who join are those who cannot find other jobs. But the identical racial brain size occurs among officers who make the Army their profession as among enlisted personnel, and among women as among men.

It may also be worth calling attention to the enormous overlap in the racial distributions. Only a 4 per cent difference separated the Asian-American from the Afro-American averages in the US Army. Clearly it is highly problematic to generalize from a racial group average to any particular individual. However, because there is about a .35 correlation between brain size and intelligence test scores³, these systematic and possible causal relationships may be of great scientific interest.

Finally, it is not true to say that I was "suspended" from my university. The premier of the province called for this

action but the university refused to comply. Although my academic freedom was violated in other respects and for a term I was forced to teach by videotape, attempts to marginalize my research findings have not made and cannot make disappear the unpalatable facts of racial group differences in brain size and related variables.

J. Philippe Rushton
Department of Psychology,
University of Western Ontario,
London,
Ontario N6A 5C2,
Canada

SIR — John Maddox¹ was correct in saying that J. P. Rushton's most "marked observation" was that women's brains are, on average, 100 cm³ smaller than men's. Maddox failed to mention, however, that Rushton's result was (as Rushton acknowledged in his paper, p. 402) a confirmation of my earlier analyses⁴. I used autopsy data ($N=1,261$ adults) from Cleveland, Ohio, to show that, after correcting for differences in body size, brains of black and white men average about 100 grams heavier than do those of their female counterparts.

This remarkable result, previously either missed or ignored by biologists (because of unpalatability?) cannot be dismissed by claiming a biased sample. How could it be that, in Cleveland, deaths occur only among men with heavier brains and women with lighter brains? Rather, as Rushton's analyses confirm, this is a real phenomenon that the larger brain of men is an evolved trait that relates to those mental abilities at which men excel, that is, spatial and mathematical reasoning abilities. Recently initiated magnetic resonance imaging studies of male and female brains, in conjunction with tests of various mental abilities, are certain to further illuminate this fascinating aspect of human biology and evolution.

C. Davison Ankney
Department of Zoology,
University of Western Ontario,
London,
Ontario N6A 5B7,
Canada

1. *Nature* **358**, 187 (1992).
2. Rushton, J. P. *Intelligence* **16**, 401–413 (1992).
3. Willerman, L., Schultz, R., Rutledge, J. D. & Bigler, E. D. *Intelligence* **15**, 223–228 (1991).
4. Ankney, C. D. *Intelligence* **16**, 329–336 (1992).

SIR — Is bigger necessarily better? If the US Army administers IQ or aptitude tests to its recruits, Rushton's dataset might at least help to resolve that question.

Barry Knight
Ancient Monuments Laboratory,
English Heritage,
Fortress House, 23 Savile Row,
London W1X 1AB, UK

Popovic says OSI is wrong

SIR — In response to the articles "OSI finds Popovic guilty on *Science* paper"¹ and "More on Gallo and Popovic"², reporting that the Office of Scientific Integrity (OSI) of the National Institutes of Health (NIH) has found me guilty of scientific misconduct in connection with the 4 May 1984 *Science* paper of which I am first author³, I should like to say that OSI is wrong on all counts.

Six of the eight proposed 'findings' of scientific misconduct were related to individual entries of "ND" in Tables 1 and 2, and rest on a definition in the legend to Table 1 that I do not recall writing. OSI claims that the "NDs" misrepresented our data because the legend to Table 1 defines "ND" as "not done", whereas laboratory notebooks indicate that immunofluorescence (IF) assays actually had been performed. I used "ND", however, to mean "not determinable" when unambiguous, quantitative evaluation of a slide was not possible. If I had meant to indicate that no assay was performed, I would have used the notation "NT" for "not tested", as I had done in previous publications.

Moreover, the qualitative "+", "+/-" and "-" data in the notebooks, if reflected in the publication, would not have altered the conclusions. For example, in Table 2, recorded data for two HIV-1 isolates indicate that an assay using human serum (ET) was negative in one experiment but positive in another, parallel one. Moreover, contemporaneous assays using serum from patient Bru, kindly provided by Luc Montagnier, showed that 12 out of 14 IF tests exhibited clear-cut positivity, ranging from 35 to 90 per cent. (The results of assays using Bru serum were not included in the *Science* paper because Robert Gallo planned to publish them together with Luc Montagnier and his co-workers.)

OSI also claims that "ND" should not be used without explaining why it is impossible to provide more precise data. Scientific journals, however, often contain articles in which "ND" is used without explanation in place of quantitative information, and often no definition of "ND" is provided in the reported results. (For example, the *Journal of Immunology* instructs authors not to define "ND" because the journal adopts the "standard abbreviation" of "ND, not determined".)

OSI also claims that the entry "10 per cent" positive cells for H35 in Table 1 was inconsistent with Elizabeth Read-Connole's laboratory notebook entries and was therefore false. But OSI misinterpreted the entries. Once Read-Connole became aware of OSI's error

(through press accounts), she provided the following written statement to my attorneys and NIH officials: "My statement 'very few cells' was a comment on the number of the cells on the slide . . . In my second statement, I was noting that even though there were 'very few cells', the cells that were on the slide were *positive* for the rabbit polyclonal antisera raised against the HIV virus . . . I was never asked [by OSI] what I meant by these two statements . . ."

OSI also stated that it was improper for me to supply a specific number based on my own independent reading of the IF slides. It was my practice, however, always to read the slides to verify my technician's reading.

OSI claims that the statement in the methods section that "The concentrated fluids were first shown to contain particle-associated RT" misrepresents the experiment actually performed. Robert Gallo confirms that I did not write that sentence and OSI admits it has no idea who did; it first appeared in the last draft of the manuscript and resulted from someone else's editing. OSI holds me responsible nonetheless for failing to notice the change or its implications in galley proofs.

Finally, as NIH has acknowledged, the minor discrepancies do not affect the basic findings of the paper and, despite recent press accounts to the contrary, the review of OSI's proposed report is not yet complete. As they say in America, "It's not over until it's over."

Mikulas Popovic

9917 Holmhurst Road,
Bethesda, Maryland 20817, USA

1. *Nature* **357**, 3 (1992).
2. *Nature* **357**, 107 (1992).
3. *Science* **224**, 497-500 (1984).

Fetal tissue banks

SIR — Your News story "Researchers reject tissue banks" (*Nature* **357**, 267; 1992) quotes Yale researchers who extrapolated data I supplied to Senator Orrin Hatch and that he and Senator Edward Kennedy used in a Senate debate on fetal tissues research on 31 March 1992. Unfortunately, the researchers erred in their computation.

The data Congress received come from a study of miscarriages conducted at three Manhattan hospitals between 1977 and 1981: I examined more than 3,500 normal, well-preserved specimens up to 28 weeks gestational age, and concluded that about 7 per cent of them would have been potentially suitable for transplant research.

Readers should be aware that the Yale

research refers only to Parkinson's disease, in which interest focuses mainly on fetal brains of a developmental age of 7-12 weeks — a small fraction of the potentially useful fetal material. Journalists who omit this detail mislead the reader, but the researchers themselves compound the difficulty by understating by half even the amount of material available for Parkinson's-related transplant research. The Yale calculation, which you report, that "usable abortions occur, on the average, only 1.4 times a year at each hospital" is an error; in fact, about twice that number would be available for the limited purpose of Parkinson's research.

Our study remains the largest and most systematic inquiry so far into the pathology of miscarriages. The results indicate that enough tissue could be obtained to make the proposed tissue banks worthwhile (J. Byrne *et al. Teratology* **32**, 297-315; 1985). Just how much, and under what conditions, would be determined by the new tissue bank programme.

Not everyone regards the tissue bank issue from a political perspective and, if it were taken out of politics, many researchers would support such a programme. Research using fetal tissue holds enormous promise, not only for transplantation, but also in cancer research, in developmental biology and in AIDS research.

Moreover, we are still far from understanding the causes of miscarriages, despite their common occurrence: there were an estimated 750,000 in the United States last year. Ectopic pregnancies are on the rise, yet their causes are still obscure. Much good will come from the wider availability of fetal tissue for research. The opportunity to use the tissue bank networks for new studies of fetal loss should not be lost.

Julianne Byrne

Boyne Research Foundation,
Washington, DC 20010, USA

Growth of science

SIR — Edward J. Krowitz claims that the developed nations must aim not only for the absolute growth of their science, but also for relative growth with respect to developing countries (*Nature* **358**, 186; 1992). At present, the rich countries of the West and North publish some 20 times as many papers per capita as do the poor countries of Africa and Asia, yet Krowitz wants to see this inequality widen even more.

Terence Kealey

University of Cambridge,
Department of Clinical Chemistry,
Addenbrooke's Hospital,
Hills Road, Cambridge CB2 2QR, UK

Recognizing good work

SIR — Few people in research would argue with Maddox¹ about a lot of published research being impenetrable to readers. The question is what to do about it. I spend a lot of my time teaching scientists how to publish 'reader-friendly' papers; one of the tools I use is to get them to review well-written papers, such as Watson and Crick's classic², slices of Einstein's theory of relativity³ and so on.

The problem is that they know the work is good and they often adjust their opinions accordingly, so I recently did something different. I gave 24 scientists a copy of a paper by Pontecorvo⁴ and asked them to read it and to analyse it according to what they thought was good or poor. I gave them no other information about the paper or why I had chosen it, but they were given as much time as they wanted to come to a decision. (The main text is about a thousand words and most were done in 20 minutes.)

I then asked them to assemble along a line ranging from 'great' to 'rubbish' with the mid-point being the dividing line between when an editor should accept it or reject it. Only one person thought it should have been published, and most of the rest clustered near 'rubbish'. We had a good exchange of views from both sides but nobody elected to change camps.

Then I told them that I had chosen it because *Current Contents* had listed it as a citation classic (that is, a paper cited more than 400 times) and because it is a well written paper. They were surprised and confused, and found these facts difficult to accept.

My own interpretation (which I discussed with them at the time) is that their judgement was clouded by a mixture of inexperience and perfectionism. Perfectionism is an occupational hazard in science; scientists can easily get hooked into rejecting anything that has a demonstrable flaw (Pontecorvo writes well, but the perfect paper has yet to be written, so my group could all find plausible — to them — grounds for rejection.)

I think this shows a weakness in the way in which we educate researchers, which is not confined to Australia. Too often, students are left to use osmosis to learn how to publish, but this is turning out a proportion (most?) who cannot recognize good papers when they see them. If they cannot do that, what are they going to model their own papers on?

My suggestion is that academics and anyone else involved in the management of research should spend time with students and less experienced scientists to

analyse important papers in order to understand how the authors made their points, not just what the points were. This should be an integral part of any higher education and my concern is that too few do it.

Robert Brown

Queensland Department
of Primary Industries,
GPO Box 46, Brisbane, Australia, 4001

1. Maddox, J. *Nature* **305**, 477–478 (1983).
2. Watson, J. D. & Crick, F. H. *Nature* **171**, 737–738 (1953).
3. Einstein, A. *Relativity: the special and general theory*. (Authorised translation by Robert W. Lawson).
4. Pontecorvo, G. *Somatic Cell Genetics* **1**, 397–400 (1975).

Page charges

SIR — Leung and Robson (*Nature* **355**, 760; 1992) and Sharma (*Nature* **355**, 104; 1992) have referred to the problems faced by authors, particularly those in developing countries, when trying to meet page charges. I would like to draw your attention to a practice that essentially amounts to inertia selling of page charges.

In early 1991, *Physics Essays*, published by the University of Toronto Press, accepted an article of mine for publication. About six months later and shortly before the proofs arrived, I received an invoice for page charges based on a rate of \$55 per page.

Such charges were not mentioned in the instructions to authors nor anywhere else in the journal. Although I refused to pay, the article was eventually published. My understanding of English law is that I am not obliged to pay because the charges were not apparent when a contract was made, that is, acceptance for publication. Nevertheless, and despite my protests, I continue to receive regular demands for payment.

I would be interested to know whether any of your readers have had a similar experience.

L. A. King

27 Ivar Gardens,
Chineham, Basingstoke,
Hampshire RG7 4PN, UK

Orphan drugs

SIR — F. Allerberger and M. P. Dierich point out (*Nature* **357**, 531; 1992) that niclosamide is a new kind of orphan drug because it has become too cheap to justify marketing in Austria. Niclosamide has also been withdrawn from the market in the United Kingdom, where we have been using praziquantel instead. Praziquantel is an excellent taenicide

with only minor side effects and has the advantage of activity in cysticercosis, schistosomiasis and hydatid disease and is not very expensive. We are told that praziquantel is registered only for veterinary use in Austria — I wonder if the same applies to other veterinary drugs such as albendazole (useful for strongyloidiasis, gnathostomiasis and resistant enterobius) and ivermectin (drug of choice for onchocerciasis)? It is a fact of life that tropical diseases have often had to wait for decent chemotherapeutic agents to emerge from veterinary medicine although they afflict considerably more than 200,000 people.

Although it would be a shame if niclosamide were to be withdrawn from those who cannot afford praziquantel, perhaps our real concern should be not for orphan drugs but for those that are never even conceived through lack of commercial interest.

Tom Blanchard

Department of Clinical Sciences,
London School of Hygiene
and Tropical Medicine,
Keppel Street,
London WC1E 7HT, UK

Political views

SIR — Paul Cumming (*Nature* **357**, 432; 1992) expresses his dismay on reading a political commentary in *Nature* (**356**, 92; 1992) about the possible establishment of a free trade agreement between Canada and Mexico. We agree that a scientific journal should not be a forum for "unsustainable assertions which seem to gain credibility through repetition".

However, in presenting these arguments, Cumming contradicts himself. First, he makes a political statement about the inconvenience of this agreement between these countries. Second, he makes an unsustained assertion that Mexico is governed by a dictator — making a political statement about his views on the legitimacy of the Mexican government.

Alfredo D. Cuarón

Sub-Department of Veterinary Anatomy,
Tennis Court Road,
Cambridge CB2 1QS, UK
and
Centro de Ecología,
Universidad Nacional
Autónoma de México,
México DF 04510, México

Rafael Chávez-Rivera

Chemical Engineering Department,
Pembroke Street,
Cambridge CB2 3RA, UK
and
Dept de Ingeniería Química,
Universidad Autónoma,
Metropolitana-Iztapalapa,
México DF 09340, México

Why pebbles float to the surface

The reasons why mixtures of sand and pebbles seem to separate when shaken vigorously have been illuminated by an intriguing simulation which, among other things, suggests that only large pebbles behave in this way.

WHY is it that when you shake a pail containing, say, a mixture of sand and pebbles, the larger pebbles tend to rise to the top? In this, the Northern Hemisphere's summer, the observation must be repeated countless times each day by untrained observers, mostly infants, some just able to walk, playing on crowded beaches. It should therefore be of some comfort to them (or to those among their parents who believe that children's questions should be answered whenever possible) that an explanation of the phenomenon has now been given by three French researchers — Rémi Jullien and Paul Meakin from the University of Paris-Sud and André Pavlovich from the Saclay Centre at Gif-sur-Yvette in the Paris suburbs (*Phys. Rev. Lett.* **69**, 640; 27 July 1992).

Naturally a phenomenon of such universal interest needs a name: 'segregation by shaking'. But Jullien and his colleagues note that shaking is merely one of the several industrially important ways in which particles of different sizes are separated from each other mechanically — simple pouring will often decant the smaller particles preferentially, vibration will do the trick, and so on.

But how to model a complicated phenomenon such as the shaking of a mixture of particles of different sizes? On the face of things it is a formidable task requiring that all the particles in an assemblage should be handled separately. That is what Jullien and his colleagues have done or, rather, have arranged that their computers should do for them. The recipe is almost as simple-minded as it could be.

The bottom of the pail is represented by a flat horizontal surface. To begin with, spherical particles of different sizes are figuratively dropped on to it in random order by means of vertical trajectories also chosen randomly. The rules are that particles hitting the flat surface keep their positions, losing their momentum instantaneously. Eventually the time comes when the figuratively falling particles do not hit the flat surface itself, but fall on top of other particles, when the rules of the simulation require that they follow the steepest possible downward path until they reach the lowest point accessible from their place of impact. When that happens, the fallen particles are irreversibly incorporated into the growing pile of sand and pebbles. That accounts for the formation of what might be called the

pail's first charge.

The next step is to simulate the shaking, which is done in the simplest way possible. The trick is to list all the figurative particles in the simulation in order of their increasing height above the flat surface (where the height is that of the centre of the sphere), together with a note of their coordinates on that surface. Next, the whole assemblage is made figuratively to fall again, each particle along a trajectory defined by its original coordinates and, crucially, with the lowest particles falling first. Otherwise, the rules of the simulation remain the same; particles large or small hitting the bottom of the simulated pail or the growing pile that covers it lose their momentum once they reach the lowest position accessible from their impact point. The process is then repeated time and time again until the larger particles (the pebbles) are all lying on or near the surface of the pile.

It is interesting that the process as modelled is simply geometrical. Neither the physical energy imparted to the particles by shaking nor the relative density of the large and small particles enters into the simulation. Moreover, the rules of the simulation do not allow the weight of the larger spheres to displace smaller particles on which they rest from their positions, so that the simulation does not even provide for local static equilibrium.

Even so, what Jullien and his colleagues have done provides a simply explanation of why the larger particles seem to float to the surface on shaking. Smaller particles lying near the bottom of the pail at one iteration will inevitably lie near the bottom at the next, perhaps more compactly so. The larger spheres, on the other hand, will inevitably be dropped relatively later (because their turn to drop will be delayed until the simulation has reached the height represented by the position of their centres) and will then have to find a resting place on a substratum of smaller spheres already in place. At successive iterations they can move up but not down.

Jullien and his colleagues have carried through endless simulations. For example, they illustrate their article with the results of numerical experiments with 50,000 small spheres and 250 larger spheres whose diameters are four times as great. Sixty shakes (as defined) are enough to bring most of the larger spheres to the surface. Cross-sections through the simulated assemblage

at different stages show the larger spheres resting on smaller spheres (three-point contact is enough, under the rules, to support each of them) in such a way that there are also substantial voids beneath them. Perhaps the simple way of explaining segregation by shaking is that the ratchet that drives the larger spheres only upwards arises from the tendency of smaller spheres to occupy these voids on successive shakings.

At least roughly, it is even possible to calculate what the upward velocity of one of the larger spheres should be. If the size disparity between large and small spheres is great enough (and ideally infinite), the void beneath a large particle will be conical in shape with an angle represented by the angle of repose of the smaller spheres — the angle of the slope of a conical heap of them, for example.

Iterations of the simulation then coat the conical void with a uniform layer of the smaller spheres consisting of all those that would have been dropped before the larger spheres simply because their centres were lower down. The prediction of this calculation agrees with the results of the simulation.

The much more intriguing conclusion is that there is a kind of critical phenomenon involved in segregation by shaking. The process seems to function if the ratio of the diameters of the large and small particles is greater than a certain value, estimated at 2.8.

From markedly greater disparities of size, the efficiency of the segregation seems independent of the ratio of the two diameters, although the upward velocity of very large particles is less than that of particles that are slightly smaller. But when the radii of the large particles are less than 2.8 times the radii of the smaller particles with which they are mixed, it seems that the larger particles move upwards only by a finite distance and then, statistically, come to rest.

Jullien and his colleagues are properly eager to point out the limitations of their simulation — its neglect of friction, energy and all that. But they are surely correct to argue that a recognition of this critical phenomenon must be of some practical significance in industry. It may also divert many of those who now find themselves marooned on beaches in the Northern Hemisphere.

John Maddox

Chloride channels revisited

Chris Higgins

CYSTIC fibrosis (CF) is one of the most common serious inherited diseases among caucasians. The main clinical symptoms are associated with epithelial tissues, particularly those of the lung, pancreas and intestine. One of the challenges of CF research has been to identify the primary defect responsible for these symptoms. The characterization of the gene at fault in CF, and the demonstration that its product, the CF transmembrane conductance regulator (CFTR), is a chloride channel, have ensured that we are well on our way to an answer¹⁻³. But these elegant molecular studies are only part of the CF story. The function of CFTR now has to be understood in terms of epithelial physiology. What is the role of the CFTR chloride channel in relation to other epithelial ion channels? How does a defect in CFTR result in the many other biochemical differences between CF and non-CF cells and tissues? And, crucially, how does a defect in CFTR result in the pathophysiology of the disease? A first step towards these long-term goals is suggested in a report by Egan *et al.* on page 581 of this issue⁴.

It is now almost ten years since altered epithelial ion conductances were first associated with cystic fibrosis⁵. Epithelial cells contain many apparently independent ion channels and, with a view to pharmacological intervention, considerable effort has been invested in identifying and characterizing the specific ion channel(s) associated with CF. A few years ago a chloride channel, the so-called 'outwardly rectifying chloride channel' (ORCC) emerged as a strong candidate for the primary defect in CF⁶⁻⁸. Several independent research groups reported that this channel could not be activated by either protein kinase A or protein kinase C in membranes from CF cells (in contrast to membranes from non-CF cells where it could be activated). As the ORCC could be activated in CF membranes by 'non-physiological' procedures, for example by large depolarizing voltage pulses, it was concluded that the channel itself was present in CF cells but that its regulation was defective.

This was the situation in 1989 when the CF gene was identified. As CFTR had all the characteristics of a membrane protein, it raised the possibility that CFTR might itself be the ORCC. Subsequently CFTR was indeed shown to be a chloride channel^{2,3}. But CF research has never been that simple: the properties of the CFTR channel were shown to be entirely different from those of the

ORCC. Furthermore, it became clear that expression of CFTR does not correlate with the presence of ORCC, and vice versa⁹. As requested in *Nature* at that time¹⁰, "Will the real Cl⁻ channel please stand up?". Researchers had to conclude either that CFTR can regulate the ORCC (an unwanted complication in the early flush of excitement following the initial characterization of the CF gene product), or that earlier conclusions regarding the ORCC were 'unfortunate'. The role of the ORCC in CF was quietly forgotten.

The report by Egan and colleagues⁴ suggests that the ORCC should be resurrected. Their data support the conclusion that CFTR and the ORCC are distinct chloride channels, but the authors also present important evidence that CFTR can regulate the ORCC. Thus in cells lacking CFTR, the ORCC cannot be activated by protein kinase A; when CFTR is expressed in these cells, the ability of protein kinase A to activate the ORCC is restored. So here we have a potential reconciliation of 'pre-CF gene' data on the ORCC with the 'post-CF gene' studies on CFTR. But, more significantly, these studies provide direct evidence that CFTR can influence the activity of another ion channel. There are many, apparently independent, ion channels in epithelial cells, and both sodium and chloride conductances seem to be altered in CF. If CFTR can regulate the activity of other epithelial ion channels it is, perhaps, easier to understand how loss of CFTR, a small-conductance chloride channel, can lead to the considerable changes in trans-epithelial ion conductances that have been measured in membranes from CF cells.

How could the CFTR regulate the ORCC? Perhaps least satisfying would be by an indirect effect. It is known that the principal mutant form of CFTR does not reach the apical membrane (where CFTR is normally located) but is incorrectly glycosylated and remains in the Golgi: this may, fortuitously, influence the traffic of other proteins (perhaps the ORCC) to the apical membrane^{11,12}. The more enticing possibility is that the effect of CFTR expression on the ORCC reflects a physiological function of CFTR. This would be of considerable importance for our understanding of CF.

Several mechanisms by which CFTR might regulate the ORCC can be envisaged. It could interact directly with the ORCC protein in the apical membrane of epithelial cells (note that we do not yet know the protein responsible for the

ORCC). But as the CFTR channel is activated by protein kinase A and protein kinase A is necessary for CFTR to influence the ORCC, it seems more likely that the effect of CFTR on the ORCC is a consequence of CFTR function. CFTR is a chloride channel and the movement of chloride through CFTR could, potentially, regulate the ORCC (although it is not easy to envisage how). Alternatively, the regulation of the ORCC by CFTR might reflect a function for CFTR in addition to its channel activity; intriguingly, the closely related multidrug-resistance P-glycoprotein is bifunctional, being associated with both chloride channel and active transport activities^{13,14}. The experimental tools are now available for these options to be tested critically.

What are the consequences of these findings for CF therapy? The observation that, in most cases of CF, the mutant CFTR protein is incorrectly localized (it does not get to the apical membrane of epithelial cells) makes it difficult to contemplate restoring the defective protein by conventional pharmacological intervention; instead, effort has focused on replacing the mutant CFTR by gene or protein therapy. If, however, the pathophysiological changes in CF are, at least to some extent, a consequence of the failure to regulate other epithelial ion channels rather than of the absence of the small-conductance chloride channel *per se*, pharmacological intervention to restore regulation to these other channels becomes an option. Evidently several important questions still remain to be answered in the wake of the quantum leap in our understanding generated by the detailed analysis of the CF gene and its product over the past three years. What is, perhaps, most exciting is that studies of cystic fibrosis not only hold out promise for treatment of the disease, but also continue to provide deep and general insights into cell biology, biochemistry and physiology. □

Chris Higgins is at the ICRF Laboratories, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

1. Riordan, J. R. *et al.* *Science* **245**, 1066-1073 (1989).
2. Anderson, M. P. *et al.* *Science* **251**, 679-682 (1991).
3. Bear, C. E. *et al.* *Cell* **68**, 809-818 (1992).
4. Egan, M. *et al.* *Nature* **358**, 581-584 (1992).
5. Quinton, P. M. *Nature* **301**, 421-422 (1983).
6. Schoumacher, R. A. *et al.* *Nature* **330**, 752-754 (1987).
7. Li, M. *et al.* *Nature* **331**, 358-360 (1988).
8. Hwang, T.-C. *et al.* *Science* **244**, 1351-1353 (1989).
9. Ward, C. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 5277-5281 (1991).
10. Frizzell, R. A. & Claff, W. H. *Nature* **350**, 277-278 (1991).
11. Cheng, S. H. *et al.* *Cell* **66**, 1027-1036 (1991).
12. Bradbury, N. A. *et al.* *Science* **256**, 530-532 (1992).
13. Valverde, M. A. *et al.* *Nature* **355**, 830-833 (1992).
14. Gill, D. R. *et al.* *Cell* (in the press).

Trapped in mid-air

Sean Washburn

If one tosses a tennis ball into a silo (Fig. 1), the ball will remain confined there in the potential energy well. If one tosses the ball so that it passes over the top of the silo, nothing much happens — the well has no effect on the trajectory of the ball. If instead one plays at the microscopic level, the results of the game can be very different. In a new experiment, described on page 565 of this issue¹, Capasso and co-workers have observed the tennis ball (an electron in a

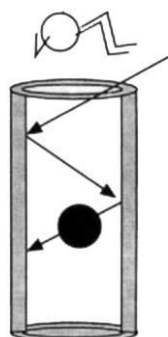


FIG. 1 Classical particles trapped inside a potential well and at rest above the top of the well.

specially designed semiconductor heterostructure), in effect, caught floating in mid-air above the top of the silo.

Quantum-mechanical objects do not hold to the standards set by tennis balls. Electrons tunnel adroitly in and out of potential wells, which accounts, for example, for nuclear α -decay. It also is the mechanism that allows the operation of the resonant tunnelling transistors^{2,3}. A tennis ball, in contrast, never tunnels through the wall of the silo. Much of the work in modern condensed-matter science involves creating objects (necessarily macroscopic ones) that mimic the microscopic physics that was formerly the exclusive domain of atomic and nuclear physics.

Some years ago⁴ carefully tailored experiments showed that, at least in the intermediate size range (with many particles but not quite enough to make a tennis ball), quantum mechanics holds: in spite of everyday experience of tossing tennis balls at silo walls, macroscopic objects do tunnel. By definition the macroscopic object contains many elementary particles, and creating a complex of particles that obey the simple equation of motion (such as the Schrödinger equation) that leads to tunnelling of a single particle takes some work — but it can be accomplished⁵. Such macroscopic tunnelling systems are rare (because the motion of all of the carriers must be intimately linked into a single macroscopic variable), which is one reason that everyday experience causes

confusion when one studies the quantum world.

The important feature of quantum mechanics in the macroscopic domain is the effect of the environment on the system being studied. Typically the system must be cooled to very low temperatures, below 1 K, so that thermal fluctuations (random noise) in the surroundings do not smear any signatures of quantum mechanics to the point that they are unrecognizable. In the case of the quantum tunnelling, the mechanism is delicate and very low temperatures are required, but the environment actually increases the tunnelling rate. Other solid-state systems exhibit both simple and subtle quantum effects⁶ — even when the carrier motion is not linked, and the carriers respond individually to the environment. Moreover, materials science is now so sophisticated that one can create an environment that enhances a particular quantum effect, and that is what Capasso *et al.* have done¹.

In the simple quantum-mechanical system (Fig. 2a) comprising equidistant wells, the wells support bound states which lie below the energy required for free motion of the particle. If an electron falls into the potential well, it resides at the energy set by solving the Schrödinger equation with a wavefunction that oscillates (typically) within the well and decays (weakens) rapidly outside of the well. This is quite reasonable — the oscillations in the well correspond to the particle bouncing back and forth between the walls, and the strong decay indicates that it is unlikely to escape. Above the top of the well, the wavefunctions oscillate throughout space and never decay. Such extended states correspond to tossing the ball across the top of the silo.

When there are many wells at regular intervals, the extended levels form bands of available energy states. It is known from chemistry and atomic physics that electrons can be poised in metastable states above a potential well for long durations. Such bound states lead to auto-ionization (or 'predissociation') and to the Auger

process in semiconductors, and are related to the crossing of energy levels^{7,8}.

If one considers a system of regularly spaced particles such as a crystal lattice, with one impurity, one finds that the charge density at the impurity is unusual. Members of the uniform lattice share the electrons and their energy equally, but the impurity provides a potential well, and it accrues charge trivially if it brings its own bound state into which the charge can fall. Alternatively, the collective reflections of electrons from the lattice at large can lead to a new bound state. The lattice environment gives the binding energy to the impurity. Similar physics appears in optics, where it is called Fabry-Pérot interferometry⁹. A clever use of this constructive interference was made by Yablonovitch *et al.*¹⁰ who constructed a 'crystal' out of machined plastic and studied 'impurities' (defects) in the artificial crystal by the way they diffracted microwaves.

Capasso *et al.*¹ have used this scheme to create a new kind of semiconductor device¹¹. The device crystal is grown with alternating layers (a superlattice) of AlInAs and GaInAs to create a stack of narrow wells that serve as mirrors; these narrow wells surround a wide well that serves as the impurity (Fig. 2b). The mirrors were separated by one-half of the electron wavelength (which depends on its energy) and reflected the electron waves so that they resonate at the wide

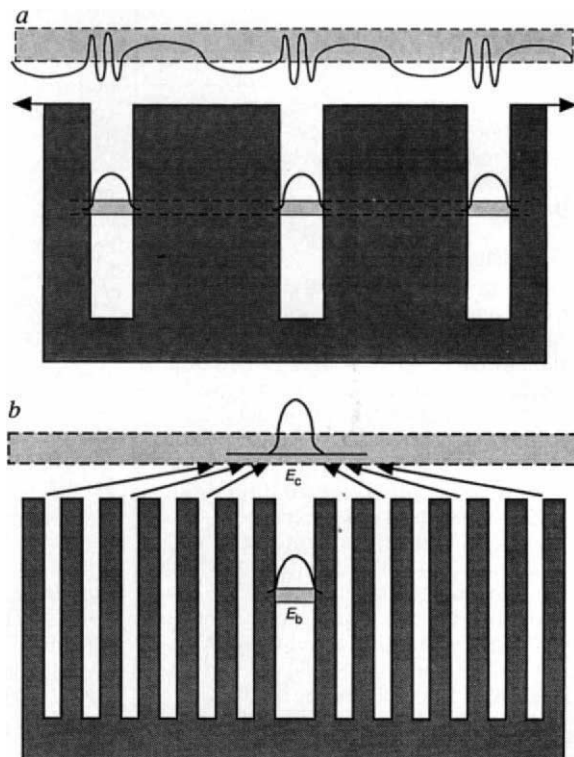


FIG. 2 a, Isolated quantum potential wells lead to bound states below the tops of the wells and to a band of extended states above the tops. But using mirrors, Capasso *et al.*¹ have created a bound state above the well (b).

RÉSUMÉ

Whipple's bacterium

It was over 80 years ago that George Whipple described a rare and then invariably fatal intestinal disease that now bears his name, and some 30 since the probable causative agent was confirmed as a bacterium. But which bacterium? In a striking example of how molecular techniques are superseding traditional methods of bacterial taxonomy, D. A. Relman *et al.* have identified the organism responsible by amplifying its 16S ribosomal RNA gene and sequencing it (*New Engl. J. Med.* **327**, 293–301; 1992). Curiously, the bacterium, dubbed *Tropheryma whippelii* seems to be akin to the actinomycetes, a group most commonly found in soil. So as well as paving the way for a new diagnostic test, the study may suggest approaches by which the hitherto intractable *T. whippelii* might be cultured for study.

Sticking point

J. H. Hoh *et al.* believe they may have observed individual hydrogen bonds using an atomic force microscope (*J. Am. chem. Soc.* **114**, 4917–4918; 1992). The microscope consists of a silicon nitride tip, mounted on a cantilever, which scans over a glass surface. Sensitive to forces as weak as 10^{-11} newtons, the tip appeared to snag on individual features on the glass surface, producing a kind of microfriction. The adhesion force was 1.2×10^{-11} newtons, although multiples of this value were also seen. The authors suggest that microscope tip stuck to the glass through hydrogen bonds — each surface carries a profusion of silanol groups — although they allow that ordered layers of water (in which the experiments were performed) will have modified the force.

Staying alive

The mass extinction at the close of the Cretaceous 65 million years ago had a far more drastic effect on land animals than those living in freshwater habitats. Aquatic creatures could have survived a temporary loss of primary productivity by relying on a food chain based on detritus feeding, according to P. M. Sheehan and D. E. Fastovsky (*Geology* **20**, 556–560; 1992). Their data come from the inventory of Uppermost Cretaceous and Lowermost Palaeocene vertebrates from eastern Montana compiled by J. D. Archibald and L. J. Bryant which, when analysed, shows that about 88% of species of land animals did not survive the end of the Cretaceous, but that some 90% of freshwater animals did. The authors emphasize that the data apply only to eastern Montana; nevertheless, the results are consistent with a variety of circumstances inferred as consequences of bolide impact.

well. The mirrors' thickness and separation were selected (and exquisitely grown) to reflect the carriers and make them interfere to form a bound state at the wide well. The energy E_c of the bound state was set by the mirror spacing to be within the continuum of extended states from the wide well.

This trapped state was observed by optically exciting electrons from a lower, bound state (E_b) of the wide well. Instead of the broad response indicative of a continuous band of above-well states, the authors observed a narrow resonance at $E_c - E_b$, just where it was expected for the particular arrangement of mirrors.

The energy and the width of the resonance could be adjusted by changing the separations and the number of mirrors. The resonance even seems to have the characteristic asymmetric shape expected for level crossings^{7,8}. These results are an elegant confirmation of an old prediction^{12,13} and a step toward more sophisticated optical-electronic circuits in which tailored energy states serve to mimic classical optics. □

Sean Washburn is in the Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27599, USA.

1. Capasso, F. *et al.* *Nature* **358**, 565–567 (1992).
2. Chang, L. L., Esaki, L. & Tsu, R. *Appl. Phys. Lett.* **24**, 593–595 (1974).
3. Mendez, E. & von Klitzing, K. (eds) *Physics and Applications of Quantum Wells and Superlattices* (Plenum, New York, 1986).
4. Voss, R. F. & Webb, R. A. *Phys. Rev. Lett.* **47**, 697–700 (1981).
5. Chakravarty, S. in *New Techniques and Ideas in Quantum Measurement Theory* (ed. Greenberger, D. M.) 25–35 (New York Acad. Sci., New York, 1986).
6. Mazenko, E. & Grinstein, G. (eds) *Directions in Condensed Matter Physics* (World Scientific, Singapore, 1986).
7. Fano, U. *Phys. Rev.* **B124**, 1866–1872 (1961).
8. von Neumann, J. & Wigner, E. *Phys. Z.* **30**, 467–470 (1929).
9. Fowles, G. R. *Introduction to Modern Optics*, 90–94 (Dover, New York, 1989).
10. Yablonovitch, E. *et al.* *Phys. Rev. Lett.* **67**, 3380–3383 (1991).
11. Stillinger, F. H. *Physica* **85B**, 270–276 (1977).
12. von Neumann, J. & Wigner, E. *Phys. Z.* **30**, 465–467 (1929).
13. Simon, B. *Commun. pure appl. Math.* **22**, 531–538 (1969).

TUBERCULOSIS

Back to a frightening future

Barry R. Bloom

LESSONS unlearned and opportunities lost have come back to haunt us¹. Tuberculosis — the largest cause of death in the world from a single infectious disease² — has been virtually ignored for 20 years and more. While TB remains a major problem in developing countries, cases have increased by 33.3% in Switzerland, 30.7% in Denmark, 28% in Italy and 11.8% in the United States in recent years. Most ominous, the emergence of multidrug-resistant strains has reduced the efficacy of treatment almost to the level of the pre-antibiotic era³. The work described by Zhang *et al.*⁴ on page 591 of this issue offers hope that molecular genetics may be one way to tackle the growing problem of drug-resistant tuberculosis. This paper reports the first characterization of a molecular target for any antituberculosis drug, even though the earliest antibiotic effective against tuberculosis, streptomycin, was introduced 40 years ago.

The drug, isoniazid (isonicotinic acid hydrazide, INH), was first synthesized in 1912 and has been the cornerstone of all effective regimens for treatment and prophylaxis of tuberculosis since the 1950s. Zhang *et al.* selected INH-resistant mutants in a rapidly growing surrogate mycobacterium, *Mycobacterium smegmatis*, because *M. tuberculosis* grows only slowly. By screening DNA libraries from virulent *M. tuberculosis*, they identified a gene that restored sen-

sitivity to INH in the *M. smegmatis* mutants and in *Escherichia coli*, which is intrinsically resistant to the drug. This gene (*katG*) encodes a catalase-peroxidase enzyme. The authors were keenly aware that clinical INH-resistant isolates often lack catalase activity, so they used DNA probes for *katG* to show that a subset of INH-resistant isolates from TB patients had deletions in this catalase-peroxidase gene. They suggest that the enzyme encoded by *katG* may chemically convert INH to a biologically active form. One hope is that it will be possible to synthesize the active INH metabolite, which, if it is stable, might then be useful in treatment. But analogues of INH are unlikely to be effective if the catalase-peroxidase gene has been deleted, so it is now important to search for the target(s) of the activated form of INH, a strategy that could lead to the development of new drugs. Because with current technology it takes between two and eight weeks to isolate *M. tuberculosis* from clinical samples, and between one and thirteen weeks to establish their sensitivity to antibiotics⁵, to have specific DNA probes for the critical mutant targets of the major drugs used against tuberculosis would be a great advantage for the molecular assessment of patterns of drug resistance. Early application of effective treatment is the key to curing and blocking transmission of this disease. ►

To what can the rise of drug resistance in *M. tuberculosis* in many parts of the world be attributed? Historically tuberculosis is a socially sensitive disease⁶ which flourishes under the overcrowded conditions found among the poor in developing and the poor and homeless in — the irony of the phrase is painful — the developed countries. Because after a few weeks of treatment patients feel better, they often take their medication irregularly or not at all until they relapse — a perfect strategy for developing drug resistance. The alarmingly heightened susceptibility to TB of individuals infected with human immunodeficiency virus has added a new dimension to the problem. One critical question is whether higher frequencies of drug-resistant variants emerge as a consequence of immunodeficiency, and the diminution of the immune response as a major selective force against the pathogen.

As demographers and epidemiologists peer into their crystal computers and ponder on the world in the year 2015, they foresee major changes. The trend first perceived in Europe and currently underway in the advanced developing countries is predicted, in time, to affect all developing countries. This Demographic Transition translates into more people living longer, population ageing and urbanization^{7,8}. In developing countries there is often a period in which death rates decline before a decrease in fertility rates, resulting in a bulge in the population. The health correlate to the Demographic Transition is the Epidemiological Transition, which is characterized by increased life expectancy, increased prevalence of chronic degenerative diseases and decreased mortality, particularly from infectious diseases⁹⁻¹². Two health problems, then, have to be addressed — the triad of infectious diseases, malnutrition and population pressure; and chronic diseases such as cardiovascular disease and cancer. The industrialized countries passed through the Epidemiological Transition only gradually, but, in the words of Foege and Henderson¹³, "The developing countries will not have the luxury of dealing with the two kinds of problems sequentially. For the remainder of this century, the developing countries will be dealing with both simultaneously". And in these countries, we should remind ourselves, live three-quarters of the human population of the planet.

One policy implication of health transition thinking is the reallocation of scarce health resources to the area where the new action is likely to be, namely to chronic illness, with a de-emphasis on infectious diseases. The recent AIDS, cholera and dengue epidemics, however, indicate that such a policy decision may

Hydra and the homeobox

THE more people look for homeobox genes, the more such genes turn up. For instance, A. Miles and D. J. Miller have isolated a homeobox gene from the coral *Acropora formosa* that bears sequence similarities to the *Drosophila* pair-rule gene *even-skipped* (*Proc. R. Soc. Lond.* B248, 159–161; 1992). Now, no fewer than five homeobox-class genes are reported in multi-headed mutants of another cnidarian, the hydra *Chlorohydra viridissima* (M. Schummer *et al.* *EMBO J.* 11, 1815–1823; 1992). Intriguingly, at least two of the genes are sequentially expressed in the regenerating head end of the animal, and in the same order as their cognates in *Drosophila* and mice. The observations imply that even in these simple, non-segmented animals (which lack a mesoderm), homeobox-class genes are involved in the specification of anterior–posterior polarity. The hydra genes are dubbed *cnx1*, 2, 3 and 4 (short for cnidarian homeobox) and *msh*, and the homeoboxes of the first two show distinct resemblances to *labial/Hox-1.6* and *Deformed/Hox-1.4* respectively. Just as *Hox-1.6*-family members are expressed more anteriorly than those of the *Hox-1.4* family, the expression of *cnx-1* peaks before that of *cnx-2* during hydra head regeneration. *cnx-3* is similar to *Distal-less*, and *msh* is similar to the *msh/Hox-7* gene family. Schummer *et al.* point out that the ready regeneration of the mutant multi-headed hydra makes an excellent model system for testing ideas about head development in higher animals.



Getting a head — regenerating hydra after 30 hours, after peak expression of *cnx2* and before peak expression of *cnx3*.

be premature; we remain vulnerable to new and emerging infectious diseases¹⁴. The paper by Zhang *et al.* reminds us that there is also a constant threat from mutations of infectious pathogens which confer resistance to previously effective drugs and to the immunological surveillance that we take for granted.

This is unlikely to be a problem peculiar to tuberculosis. *Streptococcus pneumoniae*, which has no β -lactamase, the common enzyme for degrading penicillin-like drugs, has found another way to resist destruction by these drugs¹⁵. And in the case of *Staphylococcus aureus*, a major source of hospital infections which mutated to become resistant to methicillin and was held at bay for thirty years only by vancomycin, new isolates resistant to this drug have been

reported¹⁶. Malarial parasites have emerged in Asia with multiple resistance to a variety of antimalarial drugs (P. de Raadt and T. Godal, personal communication). As in the case of multidrug-resistant *M. tuberculosis*, there are few, if any, useful back-up drugs in the pipeline.

Without a significant effort to control the appropriate use of, and compliance with, existing antibiotic regimens, and increased support for basic research and development of new drugs and vaccines, we may be working our way back to a frightening future. □

Barry R. Bloom is at the Howard Hughes Medical Institute, Albert Einstein College of Medicine, Bronx, New York 10461, USA.

1. Bloom, B. R. & Murray, C. J. L. *Science* (in the press).
2. Murray, C. J. L., Styblo, K. & Rouillon, A. in *Disease Control Priorities in Developing Countries* (eds Jamison, D. T. & Mosley, W. H.) (Population, Health and Nutrition Division, The World Bank, Washington DC, 1992).
3. Snider, D. E. & Roper, W. L. *New Engl. J. Med.* **326**, 703–705 (1992).
4. Zhang, Y., Heym, B., Allen, B., Young, D. & Cole, S. *Nature* **358**, 591–593 (1992).
5. Centers for Disease Control *MMWR* **41**, 507–509 (1992).
6. Dubos, R. & Dubos, J. *The White Plague* (Rutgers Univ. Press, New Brunswick, New Jersey, 1987).
7. Notestein, F. W. *Proc. 8th Int. Conf. agric. Economists*, London (Oxford Univ. Press, 1953).
8. Coale, A. J. & Watkins, S. C. (eds) *The Decline of Fertility in Europe* (Princeton Univ. Press, 1986).
9. Omran, A. R. *Milbank Quarterly* **49**, 509–538 (1971).
10. Olshansky, S. J. & Ault, A. B. *Milbank Quarterly* **64**, 355–391 (1986).
11. Bobadilla, J. L., Frenk, J., Frejka, T., Lozano, R. & Stern, C. *Evolving Health Sector Priorities in Developing Countries* (Population, Health and Nutrition Division, The World Bank, Washington DC, in the press).
12. Frenk, J., Bobadilla, J. L., Stern, C., Frejka, T. & Lozano, R. *Health Transition Rev.* **1**, 21–38 (1991).
13. Foege, W. H. & Henderson, D. A. in *Strategies for Primary Health Care* (eds Walsh, J. A. & Warren, K. S.) 321 (Univ. of Chicago Press, 1986).
14. Lederberg, J., Shope, R. & Oakes, S. *Emerging Threats to Health* (US Natn. Acad. Sci., Washington DC, in the press).
15. Musher, D. M. *Infect. Dis. Clin. N. Am.* **5**, 509–521 (1991).
16. Johnson, A. P., Uttley, A. H., Woodford, N. & George, R. C. *Clin. Microbiol. Rev.* **3**, 289–291 (1991).

Sponges to wipe away pain

J. Mann

It is claimed that only 10 per cent of rainforest plants have been screened for drug activity, yet these have yielded numerous anticancer, antiviral and anti-inflammatory agents. How much greater must be the potential of marine organisms, which are far more diverse. The first comprehensive review of marine natural products did not appear until 1973¹, and since then there has been a veritable flood of papers reporting the identification of novel natural products, many with both weird structures and intriguing biological activity. Elucidation of the mechanisms of action of these compounds is very preliminary, but a paper by Faulkner and co-workers² on the sponge metabolites manoalide, luffariellolide and scalaradial, provides a good example of the exciting results that are beginning to emerge.

Manoalide (compound **1**, Fig. 1) and luffariellolide (**2**) are produced by sponges of the *Luffariella* family, and scalaradial (**3**) by the sponge *Cacospongia mollior*. All are potent anti-inflammatory agents³, apparently working through inhibition of phospholipase A₂ activity⁴. This ubiquitous enzyme catalyses the hydrolysis of the ester in the *sn*-2 position of phospholipids, and is implicated in the intracellular release of arachidonic acid with subsequent formation of prostaglandins and leukotrienes via the arachidonic-acid cascade⁵.

The prostaglandins are well-known mediators of pain and inflammation, and the leukotrienes are implicated in various immune reactions and in anaphylaxis. Although many experimental drugs are now available that interfere with the arachidonic-acid cascade, few control activation of phospholipase A₂.

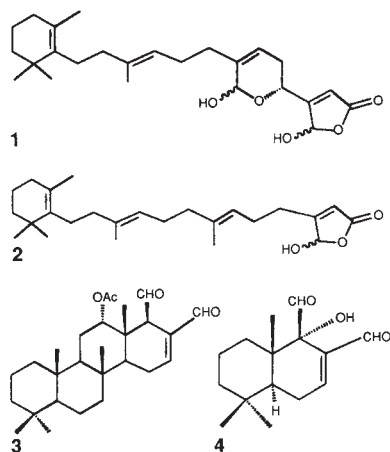
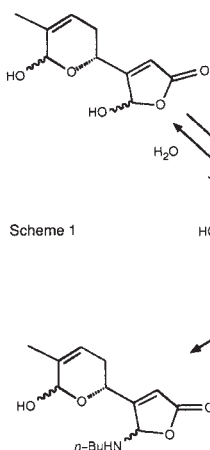


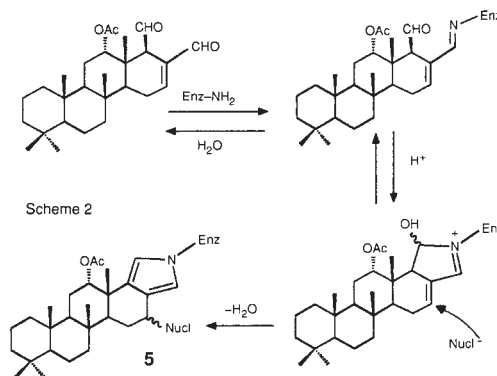
FIG. 1 The metabolites examined by Faulkner and colleagues — **1**, manoalide, **2**, luffariellolide, and **3**, scalaradial — and **4**, warburganal, an insect antifeedant⁶.

So the work with the marine natural products is clearly of considerable interest.

All three structures possess either latent aldehyde groups, CH=O, (the lactols, C-OH, of **1** and **2**) or actual aldehyde groups as in **3**, and Faulkner *et al.* demonstrate that these react with one or more lysine residues of phospholipase A₂ to form Schiff-base (imine) linkages.



Scheme 1



Scheme 2

FIG. 2 Phospholipase A₂ inactivation mechanisms. Scheme 1, the model compound reacts with the lysine-analogue *n*-butylamine ($n\text{-BuNH}_2 = \text{CH}_3\text{-CH}_2\text{-CH}_2\text{-CH}_2\text{-NH}_2$) to form a Schiff base. Scheme 2, scalaradial reacts with phospholipase A₂ (Enz-NH_2) in the presence of nucleophiles (Nucl) to form the pyrrole **5**.

Initially they worked with simple analogues of the natural products and reacted these with simple aliphatic amines (Fig. 2, scheme 1). The studies showed that Schiff-base formation is the preferred pathway. But would the enzyme studies reveal the same mode?

In the event, incubation of all three compounds with bee-venom phospholipase A₂ led to inhibition of the enzyme, but this inhibition could be reversed or prevented if the enzyme was treated with hydroxylamine (H_2NOH) at the appropriate time. This is reminiscent of work carried out with rhodopsin⁶. In the retina, retinal is covalently bound to the protein opsin via a Schiff base formed through reaction of a lysine residue and the aldehyde of retinal. In the presence of hydroxylamine (and following exposure to light), an exchange reaction occurs with the formation of retinal oxime (retinal-NOH) and regeneration of the free lysine residue of the opsin.

Faulkner *et al.* also investigated the progress of inhibition of the enzyme, and reveal that inhibition of phospholipase A₂ by scalaradial could be reversed by hydroxylamine in the early stages only, and suggest an irreversible inhibition via formation of the pyrrole (**5**) (Fig. 2, scheme 2). This is in agreement with

other model reactions tried by Cimino *et al.*⁷. Given the similarity in structure between scalaradial and other biologically interesting natural products, such as the insect antifeedant warburganal⁸ (**4**), further chemical and biological experiments are awaited with interest.

For the present, two obvious questions may be posed: can this knowledge be put to any use, and why do sponges produce anti-inflammatory compounds? It is unlikely that these sponge metabolites will ever find their way into the clinic, but this new understanding of their modes of action may stimulate the design and

synthesis of simpler structural analogues with elevated anti-inflammatory activity. Faulkner *et al.* also suggest that the reversibility of the binding to phospholipase A₂ may provide a novel means for the purification of the enzyme, through the use of affinity chromatography involving these natural products.

Finally, what of their utility to the sponges? They may simply be distasteful to predators, but it is interesting to speculate that they may bind to specific enzymes in the predator, and interfere with lipid metabolism or with some other vital part of intracellular metabolism. □

J. Mann is in the Department of Chemistry, University of Reading, PO Box 224, Whiteknights, Reading RG6 2AD, UK.

1. Scheuer, P. J. *Chemistry of Marine Natural Products* (Academic, New York, 1973).
2. Potts, B. C. M., Faulkner, D. J., de Carvalho, M. S. & Jacobs, R. S. *J. Am. chem. Soc.* **114**, 5093–5100 (1992).
3. Jacobs, R. S., Culver, P., Langdon, R., O'Brien, T. & White, S. *Tetrahedron* **41**, 981–984 (1985).
4. Glaser, K. B., Vedvick, T. S. & Jacobs, R. S. *Biochem. Pharmacol.* **37**, 3639–3646 (1988).
5. Flower, R. J. & Blackwell, G. J. *Biochem. Pharmacol.* **25**, 285–291 (1976).
6. Akhtar, M., Blasse, P. T. & Dewhurst, P. B. *Biochem. J.* **110**, 693–702 (1968).
7. Cimino, G., Sodano, G. & Spinella, A. *Tetrahedron* **43**, 5401–5410 (1987).
8. Kubo, I. *et al. J. C. S. Chem. Commun.* 1013–1014 (1976).

The target sets the tempo

Henry R. Bourne and Lubert Stryer

GRACEFUL pirouettes, exquisite *pas de trois*, tempo. . . torpid.

So goes a critique of the now familiar signalling dance of heterotrimeric G proteins (choreographed in *a* in the figure). The prima ballerina, a G-protein α -subunit, cycles between active and inactive states, depending on whether she binds GTP or GDP. Activation is swift and lissome — in a few milliseconds, hormone-bound receptor (R^* in the figure) cooperates with the G-protein $\beta\gamma$ subunit to interact with α -GDP and promote replacement of the bound nucleotide by GTP (for review, see ref. 1). By comparison, the turn-off step appears slow indeed. *In vitro*, α -GTP can take 10–60 seconds to hydrolyse its GTP and return to the inactive state (reaction 2 in the figure)^{2,3}. This languid pace cannot match the stride of many physiological responses mediated by G proteins. Some turn off in a second or less, as exemplified by the electrical response of retinal rod cells to single photons⁴ and the damping of heart rate by acetylcholine⁵.

A paper in this week's *Cell*⁶ and another in *Nature*⁷ in June help to resolve this seeming paradox. The effector molecule itself (E in the figure), the target regulated by α -GTP, can dramatically quicken the tempo by accelerating the intrinsic GTPase activity of α ; that is, GTP hydrolysis by α -GTP complexed with its effector is much faster than that of α -GTP alone (reaction 4 is faster than reaction 2 in panel *a* of the figure). Effectors thus act as GTPase-activating proteins (GAPs), analogous to GAPs for the proto-oncogene product p21^{ras} and other monomeric small GTPases.

G_q , a recently discovered member of the G protein family, mediates secretory and contractile responses of many cells to noradrenaline and acetylcholine (summarized in ref. 6). The effector stimulated by α_q -GTP is phospholipase C, which converts phosphatidylinositol 4,5-bisphosphate into two second messengers, inositol 1,4,5-trisphosphate (which releases Ca^{2+} from intracellular stores) and diacylglycerol (which activates protein kinase C). In the experiments reported in *Cell*⁶, Bernstein *et al.* added recombinant phospholipase C- $\beta 1$ to lipid vesicles containing biochemically pure G_q and recombinant muscarinic m1-acetylcholine receptors. Measurements of steady-state GTP hydrolysis and of single turnover times gave essentially the same result: phospholipase C- $\beta 1$ dramatically increased the rate of GTP hydrolysis, by as much as 50-fold, shortening the half-life of α_q -GTP from 52 seconds to 1 second or less. The faster

turn-off rate is much closer to that required for regulating physiological processes such as contraction of smooth muscle in blood vessels. The use of biochemically pure reagents leaves no doubt that both the GAP activity and the ability to generate second messengers reside in the phospholipase C- $\beta 1$ polypeptide itself.

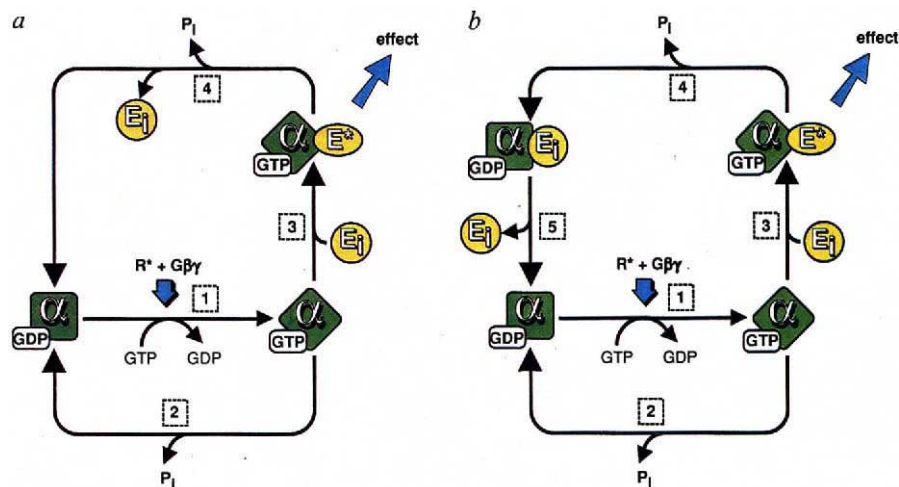
The target controls the GTPase rate in visual excitation too. First, a few words about the dancers and an enigma: absorption of a photon triggers an enzymatic cascade culminating in the hydrolysis of cyclic GMP. The decrease in cGMP closes membrane channels, thereby hyperpolarizing the plasma membrane of the retinal rod cell (reviewed in ref. 8). Photoexcited rhodopsin (R^*) catalyses the activation of transducin (G_t), which then switches on a phosphodiesterase (PDE) that hydrolyses cGMP. In the dark, the catalytic α - and β -subunits of the $\alpha\beta\gamma_2$ PDE holoenzyme are held in check by the inhibitory γ -subunits. α_t -GTP activates PDE by reversing this inhibitory constraint. In dilute membrane suspensions, α_t -GTP activates PDE by pulling γ away from the catalytic subunits^{9,10}. Under physiological conditions, the complex of α_t -GTP and PDE- γ remains connected to activated PDE- $\alpha\beta$ ¹¹. Hydrolysis of bound GTP by α_t then deactivates the PDE and the dark state is restored when cGMP is replenished by guanylyl cyclase. Excitation by a single photon takes a second, and recovery is nearly complete three seconds later⁴. In dilute membrane suspensions, however, GTP hydrolysis by α_t and turn-off of PDE activity take 30 seconds or so¹² — much

too slow for the physiological response.

The *Nature* paper by Arshavsky and Bownds⁷ helps resolve this enigma. Rod outer segments were illuminated to activate R^* and retain G_t , and washed to remove PDE. The large excess of R^* and G_t relative to substrate (radioactive GTP) ensured that single turnover times were measured. The addition of PDE holoenzyme accelerated the hydrolysis of GTP fourfold, decreasing the half-life of α_t -GTP from about 20 seconds to about five. The half-life was shorter when concentrated suspensions were used, suggesting that the hydrolytic rate *in vivo* is sufficiently fast to account for the rapid restoration of the dark state.

A second intriguing finding is that a high concentration of cGMP partially blocked the acceleration of GTP hydrolysis. Which PDE subunits stimulate the GTPase activity, and which mediate the modulatory action of cGMP? Recombinant PDE- γ had the same stimulatory effect as did the PDE holoenzyme. The GTPase-accelerating effect of PDE- γ could not be suppressed by cGMP, however. Thus it appears that PDE- $\alpha\beta$ acts as a cGMP sensor, modulating the GAP activity of PDE- γ . The duration of PDE activity will therefore depend on the cGMP concentration, an elegant servomechanism that may shape and control the time course of the photoresponse. By analogy, does phosphatidylinositol 4,5-bisphosphate, the substrate of phospholipase C, modulate the GTPase activity of α_q ?

(An aside. It should be noted that the cGMP effect was *not* observed by another laboratory using a different approach¹³: a heat signal in microcalorimetry experiments indicated a rapid rate of hydrolysis of α_t -bound GTP, even faster than that observed by Arshavsky and Bownds; similar calorimetric measurements showed light-



G-protein α -subunits cycle between inactive (square) and active (diamond) forms, bound to GDP and GTP respectively. The effector target also cycles between inactive (E_i) and active (E^*) forms, as indicated. In both panels, the upper and lower cycles show conversion of α -GTP to α -GDP with and without effector interaction, respectively.

Turning off GTP signals

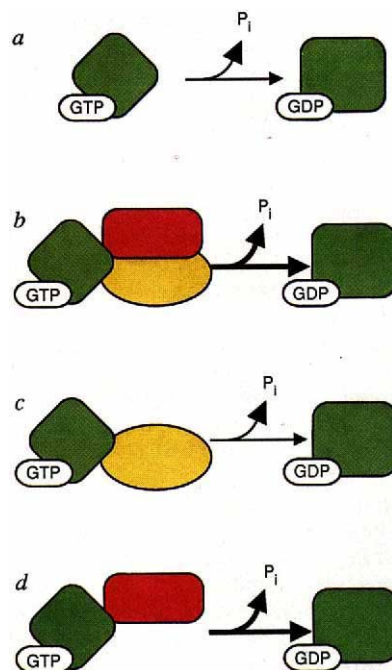
THE discovery of GAP activities in targets of the heterotrimeric G proteins brings them into step with their cousins in the GTPase superfamily¹, the Ras-like monomeric GTPases and the initiation and elongation factors of ribosomal protein synthesis. Of these, bacterial elongation factor EF-Tu has been subjected to the most intensive scrutiny. Aminoacyl transfer RNA (aa-tRNA) binds selectively to Tu-GTP, but does not accelerate its very slow intrinsic GTPase activity. Through binding of its anticodon to the mRNA-programmed ribosome, however, aa-tRNA does position Tu-GTP for interaction with a GAP located in the ribosome's 50S subunit; very quickly (within a few milliseconds), the ribosomal GAP triggers hydrolysis of Tu-bound GTP.

It is instructive to compare protein complexes that regulate GTPase activities of different members of the superfamily. Each GTPase in its unadorned G-GTP state (*a* in the figure) hydrolyses GTP relatively slowly (tens of seconds for α -GTP, many minutes for Tu-GTP). In most cases (*b* in the figure) the physiologically relevant and much faster turn-off switch is thrown when G-GTP forms a complex with other components — α_q + phospholipase C (PLC)- $\beta 1$, α_i -GTP + PDE- $\alpha\beta$ + PDE- γ , or Tu-GTP + aa-tRNA + ribosome. Just as the γ -subunit of the PDE holoenzyme can suffice to stimulate the GTPase of α_i , ribosomes by themselves (at high concentrations) can stimulate GTP hydrolysis by EF-Tu. Although neither aa-tRNA nor PDE- $\alpha\beta$ stimulates GTP hydrolysis, both of these non-GAP elements serve to regulate the GAP — by inhibiting GAP in the presence of cGMP (PDE- $\alpha\beta$) or by guiding Tu-GTP to the GAP (aa-tRNA).

Despite these and other differences, the prevailing theme is the same: although the GAP by itself can act on G-GTP, physiological turn-off of G-GTP

occurs when it binds simultaneously to its target molecule and to its GAP. Indeed PLC- $\beta 1$, a single molecule, acts as both the effector target and the GAP for α_q -GTP.

Two other models merit consideration. First (as in *c* in the figure), the



target is not associated with a GAP activity. The example of aa-tRNA and Tu-GTP shows that a target molecule may bind to a GTPase without accelerating hydrolysis of bound GTP. Some of the many isozymes of hormone-sensitive adenylyl cyclase may belong to this category. In fact, the rates at which adenylyl cyclase activities decay after hormonal stimulation^{19,20} agree rather well with measured rates of GTP hydrolysis by pure α_s ²¹.

The second model (*d* in the figure)

involves a putative 'killer-GAP' that is not associated with an effector target. Although they normally function as GAPs in association with specific target molecules, PDE- γ and the ribosome show that it is possible to deactivate the GTP-bound form of a protein in the absence of its target. At least two proteins, Ras-GAP and neurofibromin, can trigger the GTPase activity of p21^{ras}, and GAPs have been identified for many other members of the expanding family of monomeric small GTPases (for review, see ref. 22). Are these killer-GAPs or do they also act as (or in association with) effector targets? Despite generating a number of ingenious experiments, the debate over these alternatives²²⁻²⁴ has come to no firm conclusion, for a very simple reason: we are in no position to determine the relation between GAPs and effectors, because no example of the latter is in hand. Despite the many well-studied cellular consequences of activating p21^{ras}, we cannot point to a single molecule (or even a biochemical activity) that is directly regulated by the GTP-bound form of this GTPase (or of other Ras-like proteins). To speak plainly, we do not even know whether α -GTP or EF-Tu is the appropriate model for the action of p21^{ras} — that is, whether the protein works by regulating the biochemical activity of a target protein or by regulating the assembly of separate components into a macromolecular complex.

Despite our ignorance, the discovery of GAP activities associated with effector targets of G-protein α -subunits furnishes a persuasive analogy. The new observations strongly support the view that some of the GAPs for p21^{ras} and its siblings are effectors (by analogy with PLC- $\beta 1$ and α_q), whereas others (by analogy with PDE- γ and α_i) act in association with effector target proteins. The likelihood that p21^{ras} or other monomeric GTPases associate only with killer-GAPs now appears to be vanishingly small. **H.R.B. & L.S.**

stimulated PDE activity turning off at the same rapid rate, even in the presence of a high concentration of cGMP. Because both sets of results are convincing and internally consistent, it is not clear why cGMP should slow GTPase activity in one case⁷ but fail to prolong PDE activity in the other¹³.)

The exciting discovery that effectors of G proteins also act as GAPs for G proteins carries broad implications for understanding other proteins in the GTPase superfamily, including the ever enigmatic p21^{ras} (see box). This discovery also forces us to modify the prevailing model of how signalling G proteins work. In that model, the slow turn-off of α -GTP, mediated by its intrinsic GTPase

activity, seemed to account for the ability of G proteins to amplify hormone signals. Now we realize that fixing the spotlight on the prima ballerina gave predicted turn-off rates that were far too slow for physiology — a difficulty that is neatly solved by converting the turn-off into a *pas de deux*. Association of GAP activities with effectors preserves amplification of signals, but to a degree determined by the effector rather than by the G-protein α -subunit. Furthermore, the GAP activity can itself be regulated and it may vary among different effectors, thus further enhancing the versatility and flexibility of signals mediated by G proteins. It would not be surprising to learn, for example, that some of the many

recently discovered adenylyl cyclase isozymes (see ref. 14 for summary) differ primarily in their abilities to act as GAPs for the α -subunit of G_s, the stimulatory regulator of adenylyl cyclase.

Finally, we should recognize that sometimes extinction of a G-protein-mediated response may require more than the GAP activity of an effector. Until now we have assumed that only one α -GTP can activate each effector, so that the effector's GAP activity effectively turns off the response. What if an activated receptor generates a number of α -GTP molecules greater than the number of effectors available to turn them off? Indeed, an excess of G proteins over effectors may be in some cases a

kinetic necessity, to ensure quick turn-on of the response. In such a case, an effector-GAP can quickly turn off the first α -GTP bound to it. If the resulting α -GDP dissociated right away, however, another α -GTP would take its place, which would sustain the activated state. All of the α -GTP generated would have to be hydrolysed before the effector could return to the inactive state. Such a course of events may be quite common, as suggested by the 10-fold excess of G_i over PDE in rod cells¹⁵ and the even greater excess of G_s over adenylyl cyclase in S49 lymphoma cells^{16,17}.

To avoid the difficulty, each activated effector must turn off and remain refractory to further stimuli until the wave of α -GTP molecules disappears. Such a refractory period should begin when α -GTP is inactivated in the GAP-triggered turn-off reaction. What is the simplest way to pull off such a trick? Why not impose refractoriness by using the α -GDP produced by GAP activity to shield the effector from new α -GTP molecules? In this hypothetical scheme (*b* in the figure) reaction 5 is slow — ideally, slower than reaction 2. In fact, experiments with α_i and PDE have already demonstrated that such a shielding effect is possible. These experiments used a PDE holoenzyme activated by proteolytic removal of PDE- γ ; α_i -GDP inhibited the catalytic activity of the remaining catalytic subunits (PDE- $\alpha\beta$) and prevented re-activation by α_i -GTP¹⁸.

Although the possible usefulness of α -GDP for making effectors refractory

has not been demonstrated (or even tested) in a physiological context, the ability of α_i -GDP to bind and inactivate PDE- $\alpha\beta$ makes it plausible. It also enhances the potential importance of α -GDP in other signalling pathways. Until now investigators have considered α -GDP a sort of 'drone' conformation, used only to store α -subunits for later activation. Instead, we propose a new role for the prima ballerina in the intricate choreography of signalling G pro-

teins — by protecting her partner from other dancers, the persisting embrace of the GDP-bound ballerina forces the effector to take a timely rest before the tempo quickens once again. □

Henry R. Bourne is in the School of Medicine, University of California, San Francisco, California 94143, USA. Lubert Stryer is in the School of Medicine, Stanford University, Stanford, California 94305, USA.

RNA CATALYSIS

Evolution on fast-forward

William P. Jencks

CHEMISTRY and molecular biology are rapidly taking us backwards in time towards the origin of the self-replicating systems that ultimately gave rise to life on Earth as we know it today. The 'RNA world' that has been proposed as a beginning for life — or as a precursor to it — is becoming increasingly plausible. This framework for the start of life is largely built on the catalytic features of RNA^{1,2}, and investigators following these reactions may well find themselves tracking evolution as they discover increasingly complex and well-organized catalytic RNA systems. Two articles in *Nature*, one by Pyle, Murphy and Cech³ published last month, and the other by Pan and Uhlenbeck⁴ on page 560 of this issue, give new insight into the binding mechanisms that are responsible for the remarkable acceleration in rates brought about by non-covalent interactions between the active sites of RNA enzymes (ribozymes) and their substrates.

Pyle, Murphy and Cech³ use site-directed mutagenesis to probe the mechanism by which the RNA of a ribozyme accelerates the nucleophilic attack of a hydroxyl group of a bound guanosine molecule on the phosphodiester bond of the 5'-adenylate group of an oligonucleotide. The nonanucleotide substrate GGCCUCUA binds to a complementary hexanucleotide sequence of the ribozyme and donates the terminal adenylate to a hydroxyl group on a guanosine that is held in a specific binding site near the other end of the RNA sequence of the ribozyme. A different line of investigation is taken by Pan and Uhlenbeck⁴, who have developed a method to isolate RNA that undergoes autolytic cleavage as the result of activation by Pb^{2+} and Mg^{2+} ions⁵.

Nucleic acids, with only four bases, have a much smaller armamentarium of reactive groups to bring about catalysis compared with proteins, and the rate constants for catalysis are generally much smaller than they are for catalysis

by enzymes. But this is not a serious problem for the ribozyme, which does not need to catalyse a large number of turnovers in a short time. There is not a strong driving force for the evolution of ribozymes with large catalytic rate constants.

One surprising factor is the influence of the three-dimensional structure of the RNA catalyst in these reactions. We have been used to thinking of nucleic acids as the providers of a linear code for the synthesis of proteins and of other nucleic acids, but we now have to think about how the three-dimensional structure of RNA is responsible for its function as a catalyst.

Binding to ribozymes arises from favourable interactions of the catalytic centre with different groups on the substrate and not just from base pairing, so binding is much stronger than it would be with base pairing alone. For example, 2'-hydroxyl groups on sugars that are separated from the cleavage site by 2–3 nucleotides can provide up to 3 kilocalories of binding energy³. Protein enzymes achieve catalysis largely by using non-covalent binding between the enzyme and the transition state for reaction with substrate, so the transition state can be stabilized relative to the bound substrate⁵. It is likely that ribozymes also use this binding energy for bringing about catalysis by stabilizing the geometry of the transition state for phosphate transfer, as well as by induced approximate fit. This binding energy can bring about rate increases of 10^{12} or more by enzyme proteins⁶; it will be interesting to see if ribozymes can be as effective.

The construction of mutant ribozymes with different sequences and folded structures is an attractive route for learning more about the factors that can contribute to ribozymic catalysis. The rate acceleration brought about by these molecules must surely arise mainly from the induced intramolecularity of the reacting groups, which decreases the

1. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
2. Gilman, A. G. *A. Rev. Biochem.* **56**, 615–649 (1987).
3. Berstein, G. *et al.* *J. biol. Chem.* **267**, 8081–8088 (1992).
4. Baylor, D. A., Lamb, T. D. & Yau, K. W. *J. Physiol.* **288**, 613–634 (1979).
5. Breitwieser, G. E. & Szabo, G. *J. gen. Physiol.* **91**, 469–493 (1988).
6. Berstein, G. *et al.* *Cell* **70**, 411–418 (1992).
7. Arshavsky, V. Y. & Bownds, M. D. *Nature* **357**, 416–417 (1992).
8. Stryer, L. *J. biol. Chem.* **266**, 10711–10714 (1991).
9. Wensel, T. G. & Stryer, L. *Biochemistry* **29**, 2155–2161 (1990).
10. Chabre, M. & Deterre, P. *Eur. J. Biochem.* **179**, 255–266 (1989).
11. Clerc, A. & Bennett, N. *J. biol. Chem.* **267**, 6620–6627 (1992).
12. Leibman, P. A. & Pugh, E. N. *Curr. Topics Membr. Transport* **15**, 157–170 (1981).
13. Vuong, T. M. & Chabre, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9813–9817 (1991).
14. Gao, B. & Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10178–10182 (1991).
15. Stryer, L. *A. Rev. Neurosci.* **9**, 87–119 (1986).
16. Insel, P. A., Mahan, L. C., Motulsky, H. J., Stoolman, L. M. & Koachman, A. M. *J. biol. Chem.* **258**, 13597–13605 (1983).
17. Alousi, A. A., Jasper, J. R., Insel, P. A. & Motulsky, H. J. *FASEB J.* **5**, 2300–2303 (1991).
18. Kroll, S., Phillips, W. J. & Cerlone, R. A. *J. biol. Chem.* **264**, 4490–4497 (1989).
19. Cassel, D., Levkovitz, H. & Selinger, Z. *J. cyclic Nucleotide Res.* **3**, 393–406 (1977).
20. Landis, C. A. *et al.* *Nature* **340**, 692–696 (1989).
21. Graziano, M. P., Freissmuth, M. & Gilman, A. G. *J. biol. Chem.* **264**, 409–418 (1989).
22. Hall, A. *Cell* **69**, 389–391 (1992).
23. McCormick, F. *Cell* **56**, 5–8 (1989).
24. Hall, A. *Cell* **61**, 921–923 (1990).

large loss of entropy necessary to reach the transition state in a bimolecular reaction. Conversely, the size of the rate increase provides an estimate of the rate advantage for reactions of phosphate diesters that can result from induced intramolecularity. We shall see whether ribozymes will also be able to increase rates by induced approximation of the reacting groups to reach factors of the order of 10^4 – 10^5 times those of bimolecular reactions between reactants at one molar concentrations that are typically observed with enzymes⁷.

Pan and Uhlenbeck have developed an *in vitro* selection system to isolate an RNA molecule that uses Pb^{2+} and Mg^{2+} ions to activate cleavage of the RNA. It has been known for nearly a decade that cleavage of RNA is activated by Pb^{2+} and this ordinarily occurs by bringing about the intramolecular attack of a ribose 2'-hydroxyl group on phosphorus, which in turn displaces the 5'-hydroxyl group of the next residue and forms a cyclic 2',3'-diester^{8,9}. The remarkable property of the molecule isolated by Pan and Uhlenbeck is that it also catalyses hydrolysis of the cyclic diester specifically to give the 3'-phosphate product. The authors suggest that $PbOH^+$ is acting as a general base catalyst to abstract a proton from a ribose hydroxyl group and so facilitate its attack on the phosphate diester. This may be correct, but by itself it would not be expected to give rise to a very large increase in rate.

Metal ions can also bring about catalysis of ester cleavage by electrophilic catalysis, which activates the ester towards attack by a nucleophile, and by decreasing the acidity of water so as to generate a significant concentration of metal-bound hydroxide ion, which can serve as a powerful nucleophile at physiological pH values. Catalysis occurs because the dependence of nucleophilic reactivity on the pK_a of a hydroxylic nucleophile is very much less than the dependence of the concentration of the nucleophilic anion on the pK_a at pH values below the pK_a . Therefore a significant concentration of $Pb^{2+} \cdot OR^-$ can be generated at pH 7 which reacts very

rapidly. The ribozyme contains three Pb^{2+} ions and one of them might contribute to catalysis by interacting directly with the phosphate ester as an electrophilic catalyst to facilitate nucleophilic attack. The same mechanisms are responsible for the subsequent opening of the cyclic phosphate diester, a reaction that is also promoted by the enhanced reactivity of 5-membered cyclic phosphate diesters towards hydrolysis^{10,11}.

It is possible that hydroxyl groups or other bases of the ribozyme can act as general base catalysts that accept the hydroxyl proton of the nucleophilic hydroxyl group as it attacks the anomeric carbon atom of the ribose; this could be established by measuring the solvent iso-

tope effect of the catalysed reaction, providing it occurs during the rate-limiting step of the reaction. The importance to catalysis of the nucleophile structure could be ascertained by inserting other groups, for example $-SH$ or $-NH_2$, into the nucleophilic position. Evidently ribozymes are particularly attractive systems for investigating catalysis by macromolecules because of the simplicity of their structures and their ready manipulation by site-directed mutagenesis. □

William P. Jencks is in the Graduate Department of Biochemistry, Brandeis University, 415 South Street, Waltham, Massachusetts 02254, USA.

ROCK MECHANICS

Greasing the fault

Brian Evans

LUBRICATING two load-bearing surfaces has an immense effect on frictional slip, as anyone who has slipped on a wet floor knows. So, too, do aqueous fluids lubricate the great faults of the Earth's crust. In fact, the evolution of pore pressure during frictional sliding is of critical importance in many problems in earth sciences. In addition to influencing large earthquakes, pore fluids allow transport of minerals in the fluids along cracks and fissures to form valuable ore deposits. On page 574 of this issue¹, Blanpied *et al.* describe elegant experiments demonstrating that high pore pressures may develop along a slipping fault at elevated temperatures and pressure.

These results have special relevance to a conundrum involving the San Andreas fault in California. To the surprise of many earth scientists, geophysical measurements²⁻⁴ imply that the fault zone is both weaker than the surrounding material and weaker than predicted on the basis of laboratory friction tests — provided that the pore pressure is generated by a hydrostatic pressure gradient⁵. Several explanations for these puzzling measurements have been offered, but much recent attention has focused on the role of overpressured fluids. In fact, with the pore pressure elevated to close to the normal compressive stress, the fault could be weakened enough to reconcile the inferred relative and absolute weakness with laboratory data⁵.

Blanpied *et al.* suggest a simple way to produce overpressured fluids in the Earth. In their ex-

periments with machined rock samples, they mechanically loaded an artificial fault which was initially connected to an external water reservoir. As the experiment proceeded, the fault zone became hydraulically isolated from the reservoir, probably owing to solution-transfer processes within the cracked rock surrounding the fault. Once the fault was sealed, compaction of powdered rock raised the pore pressures, allowing slip to occur at stresses well below that expected for drained conditions. In a subsequent experiment, the frictional strength did not respond to externally imposed changes in pore pressure, verifying that the fault zone was isolated from the exterior.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

1. Zaug, A. J. & Cech, T. R. *Nucleic Acids Res.* **10**, 2823–2838 (1982).
2. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. *Cell* **35**, 849–857 (1983).
3. Pyle, A. M., Murphy, F. L. & Cech, T. R. *Nature* **358**, 123–128 (1992).
4. Pan, T. & Uhlenbeck, O. C. *Nature* **358**, 560–563 (1992).
5. Jencks, W. P. *Adv. Enzym.* **43**, 219–410 (1975).
6. Moore, S. A. & Jencks, W. P. *J. biol. Chem.* **257**, 10893–10907 (1982).
7. Jencks, W. P. *Cold Spring Harb. Symp. quant. Biol.* **LII**, 65–74 (1987).
8. Rubin, J. R. & Sundaralingam, M. *J. biomolec. struct. Dyn.* **1**, 639–646 (1983).
9. Brown, R. S., Hingerty, B. E., Dewan, J. C. & Klug, A. *Nature* **303**, 543–546 (1983).
10. Brown, D. M., Magrath, D. I. & Todd, A. R. *J. chem. Soc.* 2708–2714 (1952).
11. Dennis, E. A. & Westheimer, F. H. *J. Am. chem. Soc.* **88**, 3431–3432 (1966).

The isolation-compaction mechanism suggested in this issue is one of several plausible mechanisms. Pore pressures might also be elevated by fluid flow upwards from a deep source⁴ or by a fluid pressure gradient normal to the fault owing to the properties of the surfaces of fine-grained minerals in the fault zone⁶. But once seals have formed by any mechanism, the strength of the rocks becomes coupled to the state of the pore fluids in ways that produce exceptionally complicated behaviour. For example, diffusive surges of fluid might develop and propagate upwards from a deep, overpressured source of water, if the slipping fault has high permeability along strike, but low-permeability walls, and if permeability depends strongly on effective normal stress⁵.

Even withstanding incoming fluids, fault strength and stability might be influenced by several competing effects, each of which has a different characteristic timescale. Stress corrosion cracking in the surrounding rocks, pore fluid diffusion or dilatancy hardening along the fault zone may delay sliding instability to larger slip distances than predicted for the same fault system over a very long time⁷. Pressure-solution deformation might reduce long-term strength⁸, whereas the isolation-compaction mechanism described by Blanpied *et al.* would reduce short-term stability of sliding. But, in competition with the latter two effects, cementation, sealing and lithification could harden the fault rocks, causing the fault to strengthen, reducing permeability of the surrounding rocks, and increasing elastic stiffness of the surrounding rock^{9,10}. Along natural faults with complicated geometry and mineralogy, it is of course possible that all of these various processes occur in combination.

To predict the response of rocks to tectonic loading, it is necessary to understand something of the kinetic rates of each physical mechanism of fault weakening and strengthening. Compac-

tion of the fault rocks, for example, could occur either by fracturing¹¹ or solution-transfer processes, at rates which would depend on temperature, loading conditions, pore-fluid chemistry, mineralogy and porosity. Rates of crack healing and associated permeability reduction in the country rock will depend on the geometry and shapes of the cracks¹², as well as on environmental parameters.

When these processes are taken in

combination, a rich variety of mechanical behaviour is possible and, unfortunately, few firm experimental and geophysical constraints are available. Blanpied *et al.* have made an exciting contribution, but there is still plenty for the rest of us to look at. □

Brian Evans is in the Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

ASTRONOMY

Puzzling pulsations explained

Edward M. Sion

COMPRISING a pair of dwarf stars (one a white dwarf) that orbit one another almost twice a day, V471 Tauri has fascinated astronomers since it was discovered¹ in 1970. In 1986, mysterious pulsations with a period just over 9 minutes were found² in the binary's X-ray output, which must originate at the white dwarf — but how could not be determined. Now, using the Whole Earth Telescope (see box), Clemens *et al.*³ show that in between the X-ray pulses, the binary becomes optically bright. This evidence is just what astrophysicists needed to understand what is going on.

V471 Tauri is just about a made-to-measure astrophysical laboratory, where astronomers can study the physics of degenerate matter, the nature of stellar magnetic activity, and the origin of explosive close binary pairs. The two stars in the system are a hydrogen-rich white dwarf and a main-sequence K2 dwarf.

The former is a collapsed star at the end of its life and is made of degenerate matter; the latter is a highly active, mass-losing star with a magnetic dynamo not unlike the Sun's. Because the two stars eclipse one another during their 12½-hour orbit, and because the pair is in the Hyades star cluster, their age, original mass and present physical parameters are all well constrained.

A compelling reason for interest in the binary is that it has evidently evolved from the poorly understood 'common envelope' state: the more massive in the pair (now the white dwarf) had expanded to a red supergiant at the end of its main-sequence phase, engulfing its partner in a common envelope of gas; frictional drag then caused the two to spiral towards one another, giving the present close separation.

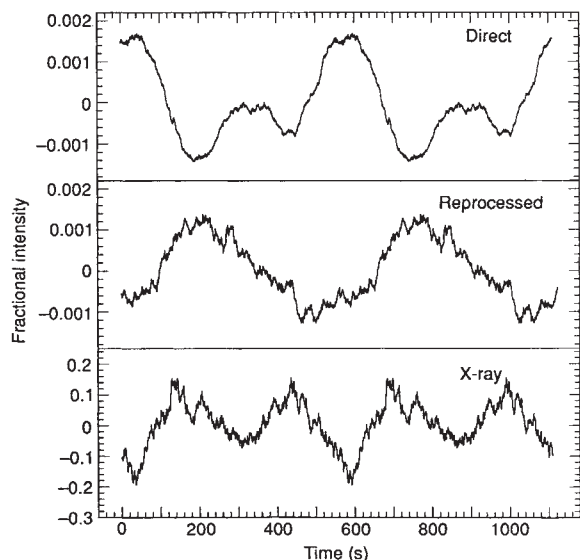
Also, the binary pair is on the verge of becoming a cataclysmic variable. With a little more shrinkage in the orbit, the atmosphere of the K2 dwarf will spill over onto the white dwarf (Roche-lobe overflow). At this point, the white dwarf's surface will accrete the excess material, fuelling the sporadic runaway thermonuclear reactions we see as classical novae. With the timescale for such a change over being astronomically small⁴ (less than a billion years), we have the prospect in V471 Tauri of viewing part of the evolution of a cataclysmic variable.

Lastly, an expanding shell of cool gas around the binary⁵ may indicate the past occurrence of a nova explosion; perhaps the white dwarf captured material from the K2 dwarf, even without Roche-lobe overflow. Chinese records of a "guest star" (nova) in the vicinity⁶ in 396 AD may be related to this.

But the crowning glory of V471 Tauri for astrophysicists was the discovery of the unprecedented 555-second pulsations in the X-ray flux², and later in optical data⁷. The eventual consensus was that the pulsations are due either to non-

THE Whole Earth Telescope (WET) is a global network of telescopes consisting of many observatories at longitudes spaced around the planet, thus allowing nearly unbroken, around-the-clock monitoring of a star's brightness. The brainchild of R.E. Nather at the University of Texas, Austin, WET is, in a sense, a single multi-mirror telescope because it is operated that way, resulting in data ten times more continuous than are obtainable from a single observatory site. To uncover and resolve all of a pulsating star's oscillation periods (most are multi-periodic) and to use their frequency spacings and amplitudes to probe the star's interior (astroseismology), the WET coverage is absolutely essential. Applied to white dwarf stars, WET has already yielded interior compositions, atmospheric layer thicknesses, rates of thermal cooling, masses, rotation rates and other fundamental physical parameters. **E.M.S.**

- Blanpied, M. L., Lockner, D. A. & Byerlee, J. D. *Nature* **358**, 574–576 (1992).
- Zoback, M. D. *et al. Science* **238**, 1105–1111 (1987).
- Mount, V. S. & Suppe, J. *Geology* **15**, 1143–1146 (1987).
- Lachenbruch, A. H. *J. geophys. Res.* **85**, 6097–6112 (1980).
- Rice, J. R. in *Fault Mechanics and Transport Properties in Rocks: A Festschrift for W. F. Brace* (eds Evans, B. & Wong, T. F.) 475–503 (Academic, London, 1992).
- Byerlee, J. *Geophys. Res. Lett.* **17**, 2109–2112 (1990).
- Rudnicki, J. *Pure appl. Geophys.* **126**, 531–524 (1988).
- Chester, F. M. & Higgs, N. G. *J. geophys. Res.* **97**, 1859–1870 (1992).
- Angevine, C. L., Turcotte, D. L. & Furnish, M. D. *Tectonics* **1**, 151–160 (1982).
- Fredrich, J. T. & Evans, B. in *Proc. 33rd US National Rock Mechanics Symposium* (ed. Wawersik, W.) (A. A. Balkema, Rotterdam, 1992).
- Wong, T.-f. in *Deformation Mechanisms, Rheology and Tectonics* (eds Knipe, R. J. & Rutter, E. H.) 111–122 (1990).
- Brantley, S. L., Evans, B., Hickman, S. H. & Crerar, D. A. *Geology* **18**, 136–139 (1989).



Direct (top), reprocessed (middle) and X-ray pulsations from V471 Tauri, for two cycles, showing phase differences. (From ref. 3.)

radial oscillations of the white dwarf, or to shadowing of the white dwarf's precessing magnetic poles as material is accreted and funnelled onto them along the field lines.

The simplest test to choose between the two models was to compare the phases of the pulsations in X-ray data and optical data. With physical pulsations modulating the star's temperature, the optical and X-ray fluctuations would be in phase. With the magnetic rotator, the pulses would be out of phase: the accreted material would block X-rays at the poles, whose diminished radiation would sweep past us as the star rotates once every 555 seconds, but would make them optically brighter, as the stellar energy must be radiated somehow.

Comparing the Whole Earth Telescope optical data with X-ray observations from the ROSAT satellite⁷, Clemens *et al.*³ come down in favour of the magnetic-rotator model, with the X-rays and optical radiation out of phase (see figure). In addition, they found optical pulses with a period of 562 seconds, attributable to reprocessed radiation from the K2 dwarf as it is periodically illuminated by its rotating white-dwarf companion. The 7-second difference in period is nicely accounted for by the slow orbit of the two stars around one another.

Where does the accreted material come from? Although there is no Roche-lobe overflow from the K2 dwarf, the star is magnetically active, like the Sun, with flares and loop structures⁸, and in 1989 we showed that it also drives a stellar wind⁹. The white dwarf's poles could be trapping material from this. It is puzzling that high-resolution far-ultraviolet spectra of V471 Tauri show none of the spectral lines that would be

expected if material is being accreted¹⁰. This will require some explanation. Astrophysicists would also like to know how large an area of the white dwarf is being shadowed, and the strength of the magnetic field funneling the wind down onto the poles.

The Hubble Space Telescope and the newly launched Extreme Ultraviolet Explorer (EUVE) satellite should help clarify matters. EUVE, in particular, should reveal the tell-tale soft-X-ray signatures of abundant elements such as C, N, O, Si and Fe when the darkened pole faces towards us, if the magnetic rotator model is correct. The degree to which the EUVE spectra change as

the white dwarf rotates will tell us something of the geometry (size, latitude) of the accreting region. The inability of the International Ultraviolet Explorer (IUE) to detect metal-line features¹⁰ places tight limits on the accretion rate of accretion of ionized species, and raises the possibility that these are batted away by the star's rotating magnetosphere while neutral helium in the K2 wind is accreted. With sufficient time resolution, EUVE spectra could elucidate this.

The ultimate prospect is of learning from V471 Tauri about the physics of diffusion in the presence of the high gravitational field of a white dwarf, about the magnetism of white dwarfs, about the physics and geometry of magnetically channelled accretion, and about the stellar wind (especially its variability) from K2 dwarfs. And if the binary truly is a strongly magnetic, pre-cataclysmic binary, then it would be the first to be identified as the immediate precursor to the well known DQ Herculis class of magnetic nova systems. □

Edward M. Sion is in the Department of Astronomy and Astrophysics, Villanova University, Villanova, Pennsylvania 19085, USA.

1. Nelson, B. & Young, A. *Publs astr. Soc. Pacif.* **82**, 699 (1970).
2. Jensen, K. *et al. Astrophys. J.* **309**, L27–L33 (1986).
3. Clemens, J. C. *et al. Astrophys. J.* **391**, 773–783 (1992).
4. Ritter, H. *Astr. Astrophys.* **169**, 139–148 (1986).
5. Sion, E. M., Bruhweiler, F. C., Mullan, D. J. & Carpenter, K. *Astrophys. J.* **341**, L17–L20 (1989).
6. Pskovskii, Y. P. *Soviet Astr. Lett.* **5**, 209–211 (1981).
7. Barstow, M. *et al. Mon. Not. R. astr. Soc.* **255**, 369–378 (1992).
8. Young, A., Skumanich, A. & Paylor, V. *Astrophys. J.* **334**, 397–411 (1988).
9. Mullan, D. J., Sion, E. M., Carpenter, K. & Bruhweiler, F. C. *Astrophys. J.* **339**, L33–L36 (1989).
10. Mullan, D. J., Shipman, H. L., Sion, E. M. & MacDonald, J. *Astrophys. J.* **374**, 707–720 (1991).

Loaded crystal

EXPLOSIVE antimony is a chemical curiosity. Made by electrolysis of antimony trichloride, it explodes violently under heat or impact, giving ordinary metallic antimony. Since a pure element cannot undergo any chemical change, Daedalus reckons that the explosion is simply an energetic change of crystal habit. He reasons that all substances, elements and compounds alike, should have such explosive crystal forms. After all, their atoms or molecules can be packed together in innumerable different ways. Some will have vastly higher energy than the normal form, but will be held in being by some energy barrier or the absence of a feasible transformation mechanism. No phase diagram will show them, but they must be there. How can they be made?

Daedalus plans to harness the crystal-modifying powers of proteins. There is a bacterial protein that nucleates the freezing of water, and a converse anti-freeze protein that stops ice crystals growing inside arctic fish. The wonderfully elaborate silica skeletons of radiolaria, and the complex shells of molluscs, are all created from simple minerals by protein-controlled crystallization.

So DREADCO's chemists are looking for such proteins. Modern bead-supported synthesis makes it easy to produce all sorts of crazy proteins to order, and the chemists are tipping them into solutions of salt, soda, and so on, to see what happens. Either by increased understanding, or by sheer blind exploration, they hope to produce ordinary substances in amazing new crystal forms, packed with enough energy to be explosive, yet stable enough to be handled safely.

The feebler varieties will come in handy for packaging. The self-opening can with an exploding metal seam, the self-destroying 'plastic explosive' bag or wrapping and the self-shattering bottle should all be widely welcomed. The apostles of planned obsolescence will rush to incorporate such materials into cars, refrigerators, hi-fi equipment, and so on. The no longer proud owner could simply light the blue touch-paper and watch the redundant item burst violently to dust and fragments. With no combustion or chemical reaction to generate flame, smoke or fumes, it could be safely set off indoors.

The more powerful explosives, perhaps energetic forms of salt or sugar, will be equally fume free. They will transform mining and tunnelling. Daedalus even has hopes of finding a high-energy ice with a melting point comfortably above room temperature. Exploding directly to steam, it would be an ideal non-polluting fuel for road transport.

David Jones

Extreme weather events

SIR — The seasonally recurring oceanic El Niño phenomenon is associated with extreme weather conditions¹. Positive phases of the tropical sea-surface temperature oscillation have been connected with unusually rainy weather over the southwestern United States, while the negative phases, usually referred to as La Niña, have been plausibly associated with extreme dry conditions over the central United States.

Since the oceanic oscillation was first connected by dynamic and thermodynamic considerations² to the atmospheric Southern Oscillation, understanding of the coupled ocean-atmosphere El Niño/Southern Oscillation (ENSO) phenomenon has increased dramatically¹, leading to hopes that these events could be predicted a year or so in advance. Statistical prediction models of ENSO³ — based on the analysis of sea-surface temperature and surface wind fields — and dynamical models⁴ can forecast events with some confidence out to 6 months or even 1 year.

These models try to forecast the detailed geographical distribution of oceanic and atmospheric variables affected by the ENSO cycle. This is a more difficult task than telling whether an El Niño or La Niña event will occur; univariate indicators of ENSO can be helpful for the latter. The Southern Oscillation index computed from time series of sea-level pressure at Tahiti and Darwin, Australia, has good diagnostic value. Minima in this index are associated with El Niño events, whereas maxima correspond to La Niña events¹.

A combination of singular-spectrum analysis⁵ and the maximum entropy method⁶ seems to hold promise for predicting the ENSO cycle 2–3 years in advance. Singular-spectrum analysis is a variant of principal component analysis applied in the time domain; it filters out variability unrelated to ENSO, separating the quasi-biennial 2–3-year variability from a lower-frequency 4–6-year El Niño–La Niña cycle; the total variance associated with ENSO combines the quasi-biennial and lower-frequency

modes⁷. The corresponding time series is shown in the figure, where it is compared with a 5-month running mean of the original Southern Oscillation index. The additional smoothness of the filtered index reflects that of the underlying temporal principal components and is crucial to the success of the forecasts using the maximum entropy method.

Measured objectively by its correlation with the observed time series in a hindcasting mode, the predictability of the principal components which carry the lower-frequency mode is between 2.5 and 3 years, a little less than the predictability of the more regular quasi-biennial mode. The figure contains a forecast based on data up to the end of February 1992. The Southern Oscillation index is expected to reach its next maximum in early 1994. The drought that affected the continental United States in the summer of 1988 has been related to the 1988–89 La Niña event and to the Southern Oscillation index maximum of January 1989⁸. The next La Niña event, if correctly predicted, could be associated with a drought over the continental United States during the second half of 1993. In addition, an El Niño event is also predicted for 1996–97, when the Southern Oscillation index, as filtered by singular-spectrum analysis, is expected to reach its next minimum.

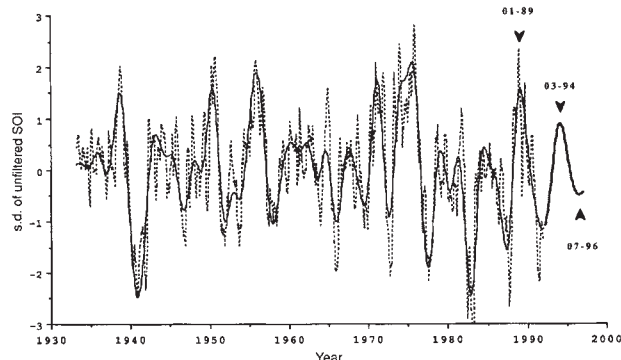
Christian L. Keppenne*

Michael Ghil*

**Department of Atmospheric Sciences,
University of California,
Los Angeles, California 90024, USA*

*†Jet Propulsion Laboratory,
California Institute of Technology,
Pasadena, California 91109, USA*

1. Philander, S. G. *El Niño, La Niña, and the Southern Oscillation* (Academic, San Diego, California, 1990).
2. Bjerknes, J. *Int. Trop. Tuna. Comm. Bull.* **5**, 219–271 (1966).
3. Barnett, T. *et al. Science* **241**, 192–196 (1988).
4. Cane, M., Zebiak, S. & Dolan, S. *Nature* **231**, 827–832 (1986).
5. Vautard, R. & Ghil, M. *Physica D* **35**, 395–424 (1989).
6. Penland, C., Ghil, M. & Weickmann K. M. *J. geophys. Res.* **96**, 22659–22671 (1991).
7. Rasmusson, E. M., Wang, X. & Ropelewski, C. F. *J. mar. Syst.* **1**, 71–96 (1990).
8. Mo, K. C., Zimmerman, J. R., Kalnay, E. & Kanamitsu, M. *Mon. Weath. Rev.* **119**, 1512–1532 (1991).



Southern Oscillation index series obtained by singular-spectrum analysis (solid, thin line); 5-month running mean of the index (dotted, thin line); and forecast obtained by applying the maximum-entropy method to the leading temporal principal components of the index (solid, thick line).

Cosmic microwave background

SIR — Fluctuations of the cosmic microwave background from the COBE satellite (discussed recently in *News and Views*¹) have been widely interpreted as providing confirmatory evidence for the Big Bang explanation of this radiation, the fluctuations being taken to represent the signatures of galaxy formation in the early Universe. Alternative explanations for the microwave background, involving thermalization, can also account for the observed energy density of the radiation and can also produce intensity fluctuations of the kind that were observed by COBE.

For such a model involving thermalization by iron whiskers of diameter $\sim 0.02 \mu\text{m}$ and length $\sim 1 \text{ mm}$, which possess a flat absorption curve over the wavelength range $\sim 0.2 \text{ mm} \approx 2 \text{ cm}$, with an average mass absorption coefficient of $1.7 \times 10^7 \text{ cm}^2 \text{ g}^{-2}$ and a smoothed-out cosmological mass density of $10^{-34} \text{ g cm}^{-3}$, the effective 'last emission' photosphere is an Earth-centred spherical surface of radius $R \approx 200 \text{ mpc}$ (refs 2–4). Fluctuations in the density distribution of iron needles in the foreground at distances nearer than 200 mpc, which are inevitable, would naturally lead to a microwave extinction which varies in intensity across the sky, but which does not alter the spectrum, leading in turn to fluctuations in the measured microwave background temperature. Some of these fluctuations could arise from clouds confined to the Galaxy, others could be of extragalactic origin.

With a reasonable assumption of density fluctuations of the order of 1 part in 10,000 occurring over inter-cluster distance scales of $L \approx 20 \text{ mpc}$, the resulting temperature fluctuations can be shown to be $\Delta T/T \approx 5 \times 10^{-6}$, as observed.

A typical angular scale of fluctuations would be $\Delta\theta \approx L/R \approx 6^\circ$, but larger fluctuations up to 90° could, of course, arise less frequently from a few local clouds. The Big Bang explanation of the latest COBE result is thus not unique, and alternative thermalization models cannot be excluded.

N. C. Wickramasinghe

*School of Mathematics,
University of Wales College Cardiff,
Senghenydd Road,
Cardiff CF2 4AG, UK*

1. Silk, J. *Nature* **356**, 741–742 (1992).
2. Hoyle, F. & Wickramasinghe, N. C. *Astrophys. Space Sci.* **147**, 245–256 (1988).
3. Wickramasinghe, N. C., Wickramasinghe, A.N. & Hoyle, F. *Astrophys. Space Sci.* (in the press).
4. Wickramasinghe, N. C. *Astrophys. Space Sci.* (in the press).

Protein function below 220 K

SIR — Rasmussen *et al.* report¹ that crystalline ribonuclease A loses its catalytic function below 220 K, because its structure becomes too rigid to undergo the small adjustments in atomic positions needed to bind substrates productively. In 1978, we found² a similar transition in the magnetic behaviour of azidemet-haemoglobin in which the haem iron is in a thermal spin equilibrium between two spin states.

Most ferric iron complexes remain in the same spin state over a wide temperature range and thus obey the Curie-Weiss law, $\chi = 1/(T - \theta)$, where χ is the magnetic susceptibility, T the absolute temperature and θ a constant; but in some complexes the separation in energy between the low (2T_2) and high (6A_1) spin states is so small that it approaches the thermal energy. Such compounds exhibit a temperature-dependent spin equilibrium. If the low-spin state is the ground state, they show a region at low temperature where they are pure low spin ($S = 1/2$) and obey the Curie-Weiss law; then a transitional region where their paramagnetic susceptibility rises with rising temperature until the high- ($S = 5/2$) and low-spin states are equally populated; and finally, a region of mixed spin where the Curie-Weiss law is obeyed again³.

The spin transition is coupled to stereochemical changes. For example, in

high-spin tetraphenylporphyrin Fe^{3+}N_3 , $\text{Fe}-\text{N}_{\text{porph}} = 2.067 \text{ \AA}$, the iron is displaced by $0.45 \text{ \AA}$ from the plane of the porphyrin and has a magnetic moment near 5.92 Bohr magnetons at 20 °C. In the low-spin pyridine complex of the same compound $\text{Fe}-\text{N}_{\text{porph}} = 1.99 \text{ \AA}$, the iron lies in the porphyrin plane, and it has a spin-only magnetic moment of 1.73 Bohr magnetons at 20 °C, which is raised somewhat by orbital contributions. When a thermal spin equilibrium exists, as in azidemet-haemoglobin, the iron oscillates between the two positions and spin states.

Addition of certain allosteric effectors has been found to raise the spin equilibrium of azidemet-haemoglobin to higher spin by switching the quaternary structure from the R (relaxed) to the T (tense) state, because the latter stretches the Fe-N bonds (see figure). In both states the paramagnetic susceptibility falls on going from room temperature to about 220 K, where it begins to follow the Curie-Weiss law, rising linearly with $1/T$. In the R-structure the magnetic

moment goes down to 2.2 Bohr magnetons before it rises again; in the T-structure the rise begins at 2.8 Bohr magnetons. In either structure the rise starts at 220 K because the globin has become too rigid for the iron atom to oscillate between two alternative positions; whatever spin equilibrium has been reached at 220 K becomes frozen in.

The loss of mobility observed by different methods in at least three proteins — ribonuclease (crystalline and suspended in supercooled liquid), haemoglobin and myoglobin⁴ in frozen solutions — suggests that 'freezing' at about 220 K may be a general property of protein structures.

M. F. Perutz

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK

1. Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. *Nature* **357**, 423–424 (1992).

2. Messana, C. *et al. Biochemistry* **17**, 3652–3662 (1978).

3. Perutz, M. F. *Rev. Biochem.* **48**, 327–386 (1979).

4. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. *Science* **254**, 1598 (1991).

Kiwi's egg size and moa

SIR — Morgan's novel speculation¹ on why kiwi have a relatively enormous egg is that kiwi became nest parasitic on moa to avoid having their ground-laid eggs gobbled by these large omnivorous ratites. Natural selection (Morgan argues) then favoured kiwi with larger eggs, as these were less easily detected in the moa's communal nest. Later, when the moa vanished and the rat arrived, kiwi survived by (presumably) learning to incubate their own eggs again, but this time in holes in the ground — the implication being that these holes afforded some protection against egg predation by these mammals.

Morgan may have sought to challenge Gould's views² on how natural selection might account for the oversized kiwi egg, but she presents an argument that runs counter to the existing data on the biology of both moa and kiwi. First, kiwi eggs have been reported in none of the more than 40 subfossil moa nests that have been found^{3,4}. Second, the suggestion that moa were omnivorous is pure speculation. Present evidence points to browse as the preferred diet of moa, along with some fruit, seeds and grass⁵. Third, moa nests contained only one, or occasionally two eggs^{3,4}. This is inconsistent with Morgan's assertion that moa had communal nests. Fourth, why kiwi should flee underground when rats arrived is bewildering. Rats and other mammalian predators are a major cause of egg loss in most New Zealand birds, but not in kiwi. In fact, the main predator of kiwi nests is the weka (*Gallinula australis*), a native rail. Nesting in a

burrow may give some protection against this avian predator, but little if any against rats.

Asking why kiwi eggs are so large in relation to the size of the bird may obscure a likely explanation by the nature of the question. Most ratites are large, and have eggs that are allometrically in proportion to their body size. If the ancestral ratites were also large (which seems a reasonable supposition), the question that should really be asked is what were the selective pressures that may have operated on kiwi to favour a reduction in body size? The availability of a niche for a forest-dwelling insectivore is one possibility.

Finally, several compelling benefits of the large, yolky eggs to kiwi hatchlings have been suggested^{6,7}, challenging Morgan's claim that the kiwi's relatively enormous egg is of no obvious advantage to the chick.

Murray Potter

Department of Ecology,
Massey University, Palmerston North,
New Zealand

John McLennan

DSIR Land Resources,
Goddard Lane, Havelock North,
New Zealand

1. Morgan, E. *Nature* **355**, 401 (1992).

2. Gould, S. J. *Bully for Brontosaurus* (Hutchinson Radius, 1991).

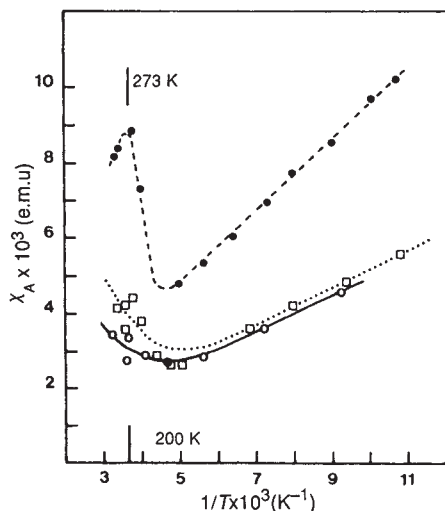
3. Hartree, W. H. *New Zealand Archaeol. Ass. Newslett.* **3**, 28 (1960).

4. Anderson, A. *Prodigious Birds* (Cambridge University Press, 1989).

5. Burrows, C. J. *et al. Rec. Canterbury Mus.* **9**, 309–336 (1981).

6. Calder, W. A. III *BioScience* **29**, 461–467 (1979).

7. Gould, S. J. *Nat. Hist.* **11**, 20–29 (1986).



Temperature dependence of paramagnetic susceptibilities of carp azidemet-haemoglobin in the R-structure without the allosteric effector inositolhexaphosphate (open circles and full line); in the T-structure with inositolhexaphosphate (solid circles and broken line); and azidemetmyoglobin (squares and dotted line). The susceptibility χ is related to the effective magnetic moment μ_e by the equation $\mu_e = 2.828\sqrt{\chi T}$.

Sobering analysis

Griffith Edwards

Children of Alcoholics: A Critical Appraisal of Theory and Research. By Kenneth J. Sher. University of Chicago Press: 1991. Pp. 226. \$28.75, £19.95.

FOR nineteenth-century Temperance activists, 'the drunkard's child' was a potent propagandist image. Tear-stained and in rags, the young victim is carrying the bottle back from the tavern to the violent father. In a once popular series of illustrations by George Cruickshank, the inebriate's decline towards madness and Bedlam is charted not only by the cat flying out of the door never to return, the hearth with no coals and the bailiff taking possession of the furniture, but also by the progressive, pitiful degradation of the children towards begging on the street.

What do we make of such images today? For most of us, those Cruickshank illustrations and their like will have lost their emotional charge. We may admire their draughtsmanship, but they are otherwise likely to be perceived by modern society only as a quaint record of a long vanished social evil.

Kenneth Sher's thoroughly modern monograph is amnesic for the nineteenth century, offering data as a substitute for etchings, and cool scientific analysis in place of emotional pleading. His book speaks, however, to exactly the same pains as the forgotten temperance tracts. Rather than the problem having gone away, the nomenclature has changed: the inebriate has become the alcoholic, while the drunkard's children have become COAs (children of alcoholics). Sher quotes epidemiological evidence to show that in the United States the lifetime prevalence for "alcohol abuse and dependence" is 25 per cent for men and 4-5 per cent for women. He also reports that in that country there may be about 6.6 million children affected by such parentage who are under 18 years of age, and 22 million aged 18 and over. The children no longer beg on the streets, but far from the problem being smugly dismissible as a phenomenon bred only by the disrupted cities and cheap liquor of the Industrial Revolution, this is a trauma that still — and on a grand scale — affects modern society. The problem has not gone away but is hidden behind curtains.

Over recent years, though, the previous and long-continued twentieth-century neglect has in the United States given way to a new social consciousness. Not only is the plight of young children who are imminently experiencing the alcoholic drinking of a parent being

given increasing attention, but a campaigning movement has developed around the supposed needs of a newly identified sector of society designated as ACOAs (adult children of alcoholics). The more strident campaigners would assure us that all ACOAs are deeply traumatized, should "get into recovery" and are universally in need of therapy. A health insurance industry that has only recently succeeded in bringing the ra-

ism in a multicausal field with complex interacting genotypic and environmental influences. Data from the US Epidemiological Catchment Area Study show, for instance, a diagnostic rate of less than 1 per cent for antisocial personality disorders among the general adult female population and a 16 per cent female lifetime prevalence for major depression. The corresponding diagnostic rates among women with drinking problems are 10 per cent for personality disorder and 31 per cent for depression. The latter figures are unlikely to be *post hoc* diagnostic artefacts.

An important variant of the genotype-environment debate arises in attempts to understand why alcoholism runs in families. A child of an alcoholic parent is between three and five times more likely to become an alcoholic than a child born

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Temperance propaganda — "The Drunkard's Children" by G. Cruickshank (1792-1878).

pacious costs for alcoholism treatment under control may be excused a little nervousness.

Sher's painstaking review amply chronicles what research reveals about the reality of the immediate traumas and the longer-term disabilities, while at the same time providing a corrective to the more lurid propagandist exaggerations. The types of harm that may be inflicted on children of alcoholics are certainly multiple. Specific damage can be done to the unborn child by the mother's drinking, either in terms of the fetal alcohol syndrome, more minor morphological aberrations or developmental delays. Much of the psychological damage is nonspecific in its manifestations, rather than constituting a specific COA or ACOA syndrome. Sher gives thoughtful attention to the difficulties that beset attempts to establish the extent of any direct contribution of parental alcohol-

to non-alcoholic parents. Despite extensive literature on family, twin, adoption and genetic-linkage research, it remains difficult to know how much of this transmission is genetic in origin, and how much is due to social learning.

Anyone with an interest in alcoholism will welcome this monograph as an outstanding review. It brings order to an area where there was previously a dominance of speculative assertion. Sher implicitly forces the large and awkward question of whether today's society is likely to generate the capacity to address the central and general issue of alcohol misuse with due, rational and sustained seriousness. In our time we have not done as well as the Victorians. □

Griffith Edwards is at the National Addiction Centre (Institute of Psychiatry), Addiction Science Building, 4 Windsor Walk, London SE5 8AF, UK.

Mary Evans

Seasoned history

Clifford M. Foust

Salt and Civilization. By S. A. M. Adshead. Macmillan/St Martin's Press: 1992. Pp. 477. £45, \$39.95.

COMMODITY history? Is that what the author of this book would have us involve ourselves in — the tracing of a single product of nature through the ages and in different economies and societies? Whatever happened to the grand view in which historical scholarship entertained large questions and was played out on a broad landscape? Well, the likes of Paul Kennedy still tempt us with the important lessons that can be learned by looking at diverse political economies over a long time span, but the lure of microhistory is equally powerful to many. The art may be in the detail.

Commodity history and other particularist approaches have striking advantages. They can help us observe specific individuals, institutions and events in the context of broad socio-economic changes, allowing, at the very least, the changes to be better illustrated, tested and understood. These approaches can also encourage a specific rather than a generalized view of comparative, international history. This kind of history is not entirely new. Since R. N. Salaman's path-breaking 1949 study *The History and Social Influence of the Potato* (Cambridge University Press), others have looked closely at sugar, cotton, tea, wine, maize, furs and even lowly rhubarb; indeed, a colleague of mine has ventured to focus on man-made products as well, such as celluloid, tin and rubber. But salt must certainly be among the most important natural commodities of all time, essential to human and animal life and, recently, a necessary part of the developing chemical industry.

Adshead divides his long study into two parts: "Salt and Society" and "Salt and the State". The first part, it turns out, concentrates on the production and distribution of salt from primitive times to modern days. Although diligent effort is made to cover equitably all major regions — the Mediterranean world, Africa, the Middle East, India and South Asia, Northern and Western Europe, Russia and the New World — there is an emphasis on China (for example, for the period 1500–1800, China gets 18 pages, Europe 11 and the Ottoman empire and Mughul India as few as 2). This is understandable, for Adshead is a reputable Sinologist, having written in 1970 the monograph *The Modernization of the Chinese Salt Administration* (Harvard

University Press). Although he strives to say what he can about salt consumption in several different societies and eras, dietary history is still not yet full enough to allow him to generalize, except by inference from overall food patterns. We do learn, nonetheless, that vegetarianism sharply limited India's consumption of salt; high per-capita use of salt has almost always been associated with meat preservation and flavouring — indeed, the spread of Islam led to greater salt use because of improved diet.

The principal techniques of salt production throughout the world have been the filtration of saline earth or ashes, followed by boiling (a procedure dependent on the availability of cheap fuels); the drilling of deep brine wells and evaporation by burning gas, coal or wood; and evaporation and condensation of natural and artificial salt seas or lakes. The last of these usually involved solar evaporation in basins or pans. The development of these techniques, Adshead argues, contributed to or reflected the general state of technical skills in each society. Sometimes salt led technical growth, sometimes it lagged behind it, depending, it is tempting to guess, on consumption demand and society's commitment to technology.

As for distribution, salt was an article of trade that from primitive times was considered as a staple. It was traded locally as an important commodity, but in some places it also formed the basis for the development of regional commercial networks, as in mediaeval Genoa and especially in Venice. Most famously, the gabelle (salt tax) of Old Regime France governed what Adshead insists was the largest European salt market. The tax, which was levied until 1790, enabled the government to cope with the "premature overpopulation" of the country and to maintain relatively high living standards.

Here we reach the second half of the book, "Salt and the State", in which Adshead deals with the involvement of governments in the production and distribution of salt and, of course, the taxation of these processes. In many although certainly not all economies, it took little imagination for rulers and bureaucrats to realize the fiscal attractions of such a socially necessary commodity. Salt production has always been highly conspicuous, making it an easy target for the tax collector's claws; and as a bulky commodity in trade, salt is not readily smuggled. Although its value was never high compared with commodities such as spices, salt remained a dependable commercial investment.

Adshead provides several examples of heavy state involvement with salt — in Venice, Old Regime France, the eastern Habsburg empire, the Ottoman empire,

India under the British Raj and, of course, China. The point is well made that administrations have rarely avoided the temptation to monopolize or otherwise manipulate this commodity for the state's fiscal ends: salt administration more often than not contributed signally to the development of fiscal, commercial and even technological institutions, and, in doing so, undoubtedly shaped the modernization of societies. However, there remains a lingering suspicion that the dominance of bureaucracies sometimes inhibited healthy technological development. And where, one wonders, is private entrepreneurship?

This is not an easy book. It is detailed, and Adshead uses lots of specialized terms and names. Although there are comparisons throughout, particularly with the Chinese scene, broad characterizations and summaries are all too rare. There are so many place names that the absence of maps might reduce one to tears. But despite these reservations, the book is an important contribution to commodity history and a laudable step towards truly comparative history. □

Clifford M. Foust is at the College of Arts and Humanities, University of Maryland at College Park, Maryland 20742-7315, USA.

States of transition

P. W. Atkins

The Chemical Bond: Structure and Dynamics. Edited by Ahmed Zewail. Academic: 1992. Pp. 313. \$49.95, £33.

FEW chemists have the privilege of seeing their ideas pervade the whole of the subject, from the static aspects of molecular structure to the dynamical details of chemical reactions. But Linus Pauling, who was born in 1901, can look back on a life well spent in contributing to the progress of understanding in the field. This volume is a record of a meeting, held last year in Pasadena, California, of some of the world's leading chemists to celebrate the impact of Pauling's work on conventional chemistry. The volume itself shows the seminal influence of the various strands of his life, for, in addition to Pauling, five of the contributors are Nobel laureates.

Pauling had the happy chance of being present when quantum mechanics burst onto the stage, and of being able to work with one of the greatest of the century's slightly undervalued physicists, Arnold Sommerfeld, when the application of

quantum mechanics to chemistry was virgin territory. It was fortunate for chemistry that Pauling was the man in place, for his extraordinary good sense about chemical matters meant that he could approach chemical problems in a sure-footed and generally sensible way.

Pauling was at the forefront of theoretical chemistry in the 1930s, when one of his main achievements was the formulation of the valence theory of chemical bonding. But this is barely mentioned in this collection of essays. Modern taste, not a little influenced by the impact of computers, views valence-bond theory as an influential but little used backwater. Although the book deals with structure, it is virtually silent on the current overwhelming domination of electronic-structure calculations based on the paradigms of molecular orbital theory. On that approach to structure, Pauling had only a waspish impact.

Nevertheless, although his early work on electronic structure is outside the mainstream of today's stunning repertoire of success, the legacy of his valence-bond theory pervades chemical thought, and few chemists would travel far without the concepts of resonance and hybridization. Dudley Herschbach, in this volume, manages to forge a link between modern thought and Pauling's influence in his idiosyncratic but interesting variable-dimensional approach to molecular-structure calculation where Lewis structures and resonance between them emerge naturally from the Schrödinger equation.

Pauling's principal lasting contribution to structure lies in his approach to the geometry of atomic arrangement; here he is a true hero, and here his upbringing in the milieu of X-ray diffraction has borne its most substantial fruit. Moreover, its intricacies still drive him on, as we see here in his pursuit of the resolution, to his satisfaction, of the existence of icosahedral quasicrystals. How pleasing it would be to think as clearly at 19 as he does at 90.

The contribution that Pauling's structural ideas made are well summarized in the essays by Max Perutz (on the importance of the hydrogen bond in physiology) and Alexander Rich (on molecular recognition between protein and nucleic acids). Perutz magnanimously acknowledges that a rule he formulated in 1986 had in fact been anticipated in a different context by Pauling half a century earlier. But he did discover something Pauling did not predict (just as others have discovered features that are contrary to what the fecund but not infallible Pauling did predict), which is the existence of hydrogen bonding in π systems (as in benzene). Rich's lengthy article is a helpful review of his field and shows just how complex the interactions

are that enable a DNA double helix to recognize a protein molecule.

The style of articles in the collection is almost as diverse as the modes of molecular recognition. Whereas Rich gives a lengthy technical survey, Francis Crick provides an essay that is a reminiscence built around Pauling's contribution to molecular biology, not the least of which was the latter's *General Chemistry*, which Crick confesses taught him the only chemistry he ever knew. This article will be of interest to those who would like to know how some (Pauling among them) took wrong roads to the helix whereas others took the right road. Crick makes a generous tribute to Pauling in ascribing to him the role of one of the principal founders of molecular biology.

The second half of the book is concerned with dynamics. Here it is harder to see the contributions that Pauling has inspired, despite the fact that his book *The Chemical Bond* has "Structure and Dynamics" as a subtitle. George Porter gives a lucid and interesting account of the emergence and application of flash-photolysis studies and a historical survey of the ever-decreasing timescale on which chemical events can be monitored. Thus we have come a long way from H. W. Melville's statement (made in 1947) that "the direct methods of physical measurement simply cannot reach these magnitudes, far less make accurate measurements in a limited period of time, for example 10^{-3} sec" to the femtosecond work that is being carried out in Imperial College in London and the California Institute of Technology. On this timescale, dynamics are snapshots of evolving structure.

John Polanyi provides a historical survey of the inception and elaboration of the concept of transition state, which is brought up to date with discussions of triangle plots and of the details that molecular beams can provide about the most intimate moments of a chemical reaction. I have already remarked that Herschbach uses his essay on chemical reaction dynamics to give an account of his variable-dimensional approach to approximation procedures, and thereby brings a different perspective to Pauling's insistence on the importance of resonance. In the bulk of his article, he also cleverly hangs his survey of molecular reaction dynamics on two other typical Pauling concepts, namely electronegativity and hybridization. The apotheosis of this climbing of Mount Resolution is the account given by the late Richard Bernstein, as compiled by Zewail, of the current spectacular work on real-time laser femtochemistry and the direct observation of transition states. Here is found the true synthesis of structure and dynamics.

Zewail is to be congratulated on bringing together such a range of contributors who, in varying degrees, show how fruitful Pauling's presence on Earth has been. The book is a fascinating summary of a goodly portion of the state of modern chemistry. □

P. W. Atkins is a fellow of Lincoln College, Oxford OX1 3DR, UK.

Tentative triffids

Martin Ingrouille

The Action Plant. By Paul Simons. Blackwell: 1992. Pp. 323. £25, \$27.95.

PLANTS are exciting — and excitable; that is the style and message of *The Action Plant*. The full range of plant behaviours, especially reversible movements, are described, but not plant growth and development. Familiar examples are the Venus fly trap (*Dionaea muscipula*), which closes the valves of its trap on its insect prey, and the 'sensitive' plant (*Mimosa pudica*), which closes its leaves almost with a shiver when touched. Less familiar are the many explosive or gentle movements of flowers and fruits. Also included are the ordinary gentle movements of plants: the opening and closing of the stomata, the pores in the surface of the plant that allow it to breathe; and cytoplasmic streaming, the movement of the cytoplasm around the cell. These movements are described in admirable detail and with excellent reference to the older and often overlooked literature, as well as to the most recent.

Paul Simons writes with freshness and clarity for the general reader. The content of the book is well researched, and there are excellent illustrations throughout. The professional botanist might quibble over some of the terminology. Fungi and slime moulds are included with green plants as if they are part of the same "Vegetable Kingdom". Chapter headings such as "Flower Power", "Bloody Plants" and "Sunbathing, Sleeping and Rhythm" might annoy some readers. But this is a small price to pay for the improved accessibility conferred by this jazziness.

Simons' main theme is the great similarity between animals and plants (hence the chapter headings "Plant Muscles" and "The Evolution of the Nervous Plant"). There is good justification for this comparison. The extent of electrical activity, including action potentials and variation potentials, and their importance in the responses of plants to touch,

wounding, extremes of temperature and even to pollination have all been demonstrated. Plants can be anaesthetized by chloroform, ether or even 1 per cent morphia. And in 1952, Loewy demonstrated the presence of actomyosin in slime moulds and proposed a theory of how actin and myosin could slide over one another to propel these organisms. This was before Hodgkin and Huxley put forward their Nobel-prizewinning model of animal muscle contraction.

But it is something of an affront to claim, as Simons does, that "plant movements are simply the manifestation of a deep-seated 'animalness' in plants". The many similarities between plants and animals are remarkable only because of our ignorance and zoocentric viewpoint. It would be more remarkable, in fact, if there was nothing similar in the cell chemistry and communication of multicellular plants and animals.

One of Simons' most interesting examples is the role of aspirin as a plant hormone. Aspirin, once extracted from willow tree (*Salix*) bark, is widespread in plants. Simons speculates that it may work in plants as it does in animals, knocking out pain-triggering prostaglandins and the closely related hormone jasmonic acid. Another mechanism is indicated by the relationship of salicylic acid to gentisic acid, the active ingredient of Ricca's factor, which is the trigger of electrical signals in the plant wound response.

Plants are exciting, however, not because of their similarities to animals but because of their differences. We are too easily seduced by the flashiness of rapid movement, but if we were able to see plants in a time-lapse film, we would see the most remarkable aspect of their behaviour: response to the environment not by movement but by growth. It is the remarkable plasticity of plant development that is thrilling. And the evolutionary consequences of this plasticity and the propensity of plants to clonal reproduction are intriguing. Some plants have lifetimes so long that they are almost immortal. Plants even challenge our concept of the cell, because plant 'cells' are connected by cytoplasmic strands forming a continuous network, a symplasm bounded by a single membrane reaching through the cell wall in fine pores, the plasmodesmata. The plant cell wall, for too long regarded as a kind of dead exoskeleton, is as much a part of the living cell as the plasmalemma.

Nevertheless, this is a very entertaining book, an easy and illuminating read. One area is not mentioned. What are the implications of the 'animalness' of plants for our eating habits? If a wave of electrical activity, reminiscent of pain and distress in animals, passes through a

lettuce leaf as we bite into it, are salads to be banned, and can anyone recommend a humane way of killing a plant? □

Martin Ingrouille is in the Department of Biology, Birkbeck College, Malet Street, London WC1E 7HX, UK.

Ain't misbelieving

Colin Howson

Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory. By John Earman. MIT Press: 1992. Pp. 265. \$35, £31.50.

SCIENTIFIC theories, even those dignified as laws, are logically underdetermined by the observational data on which they rest. A well-known philosopher's game is to produce, on request, a recipe for constructing an infinity of 'explanations' for any finite body of data inconsistent with the accepted scientific explanation. These are presumably not serious candidates for belief. But why not? Why do we regard some extrapolations of experience as more credible than others?

By the beginning of the nineteenth century, encouraged by the pioneering work of the English clergyman Thomas Bayes and by Laplace's even more remarkable extension of it, some people thought that they had the key to the answer: observational data endow hypotheses with probabilities which could (they thought) be straightforwardly computed by means of a theorem of the new mathematical theory of probability, a theorem now known, with some historical inaccuracy, as Bayes' theorem. All that remained was to do the calculations.

Unfortunately, the method by which the so-called prior probabilities in Bayes's theorem were determined, known in anglophone countries as the principle of indifference, depended on making fairly arbitrary decisions about the choice of appropriate possibility spaces. Different choices give different answers. Unsurprisingly, probabilism fell under increasing suspicion, until in the 1920s and '30s its condition was pronounced hopeless by, among others, the philosopher Karl Popper and the statistician R. A. Fisher.

But even as Fisher and Popper delivered their verdict, probabilism was finding a new and more modest role, as a theory of individuals' personal probability assignments subject to the sole constraint of consistency. The troublesome principle of indifference, or indeed any other for determining prior probabili-

ties, is arguably more than a consistency constraint and is consequently disqualified. Slimmed-down, new-model probabilism, now called 'personalist Bayesianism', or just 'Bayesianism', nevertheless still provides, Bayesians claim, as much of a positive answer to the original question as is in principle possible. The evidence they cite in support of their claim is usually some variant of the following result. Where H is a set of hypotheses about data generated from a given source, all those probability distributions that initially assign the same members of H positive probability converge, as the data accumulate, in such a way that, for almost all data sequences, they assign the true hypothesis probability one in the limit.

Do convergence theorems such as this really offer even a partial solution of the original problem? Earman's excellent new monograph on Bayesian confirmation theory contains one of the few published discussions of this difficult question to be both lucid and authoritative. Possibly its most interesting and novel feature is that it reveals how those theorems relate to recent results in what is called formal learning theory. The same high standard of discussion is maintained in the rest of Earman's book which offers, among other things, a clear and accessible survey of the attempts to develop the basic principles of Bayesianism from conditions on acceptable bets; the extent to which the theory withstands the critical assaults that have been made on it; and how, taking as an example rivals to general relativity, in which field Earman is an expert, it copes with the grand scientific theories, which are much more than merely infinite extrapolations of finite data.

Earman's investigations lead to the deflation of some of the Bayesians' claims, and the qualification of others, although overall he still rates theirs the best among extant confirmation theories. If there is any complaint about this book, it is that it has a rather narrow focus, and such topics of current interest as calibration, invariant prior probabilities, upper and lower probabilities, the claims of the Dempster-Shafer theory, and Bayesian networks are not discussed. But it is the right of an author to select his material, and the manner in which Earman addresses those issues that do concern him merits only praise. His discussion is clear, even-handed and authoritative, and, not least important in a subject increasingly prone to technicality, conducted almost everywhere in readable and lively English. □

Colin Howson is in the Department of Philosophy, London School of Economics, Houghton Street, London WC2A 2AE, UK.

Are superconductors really superconducting?

David A. Huse, Matthew P. A. Fisher & Daniel S. Fisher

The most striking difference between the behaviour of the copper oxide high-temperature superconductors and previous low-temperature type II superconductors is the much more gradual decrease in electrical resistance with temperature in the latter, in the presence of a magnetic field. This raises the question of whether a type II superconductor has strictly zero resistivity, when cooled in a magnetic field. Theoretical and experimental evidence now suggests that as the temperature is lowered, there is a sharp phase transition to a truly superconducting, impurity-dominated phase containing a disordered, frozen arrangement of magnetic flux vortices.

IN 1911, Kammerlingh-Onnes found that the electrical resistivity of mercury abruptly vanished below a temperature of 4.2 K. This historic observation of superconductivity led to the discovery of a host of related phenomena, including the expulsion of magnetic fields (the Meissner effect) and magnetic flux quantization. Attempts to understand these phenomena culminated in the phenomenological theory of Ginzburg and Landau. The microscopic theory of Bardeen, Cooper and Schrieffer (BCS) provided a justification of Ginzburg-Landau theory and explained many other properties of superconductors¹. Although most of the macroscopic properties of superconductors were understood¹ by 1970, there remained theoretical problems with understanding the apparent vanishing of the resistivity in a wide class of superconductors: the 'type II' materials. Ironically, the original and defining property of superconductors, zero resistivity, was in some ways the least well understood.

The discovery of the high-temperature (high- T_c) copper oxide superconductors in 1986 led to a resurgence of interest in superconductivity and to the development of a new class of type II materials in which the superconductivity resides on layers of CuO_2 . Although interest has focused on the microscopic 'mechanism' of superconductivity in these materials and their peculiar normal-state properties, many of their macroscopic properties are far more strikingly different from previous superconductors. In Fig. 1, the resistivity of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (known as BSCCO) is shown on a logarithmic scale for various magnetic fields as a function of temperature. In zero magnetic field, the resistivity seems to vanish abruptly at a transition temperature of $T_c \approx 90$ K, but in moderate magnetic fields of a few tesla, the resistivity is measurable down to temperatures below 20 K in spite of the expectation that it 'should' have vanished abruptly at a temperature $T_{c2}(H)$ near 80 K. Indeed, in fields greater than 5 T, BSCCO does not become a better conductor than good copper until below 30 K. These data raise fundamental questions. Is BSCCO really superconducting (with truly zero resistivity) at low temperatures in a magnetic field? More generally, do type II superconductors undergo a true phase transition to a state with strictly zero resistivity when cooled in a magnetic field, or does the resistivity just become too small to measure? Here we review recent attempts to answer these and related questions. Although these issues can be addressed entirely within the context of conventional Ginzburg-Landau theory, albeit with unconventional values of parameters, the answers require an understanding of the interplay between strong thermal fluctuations, randomly placed impurities and other crystal imperfections, and transport in disordered media. Each of these areas has seen extensive development of theory and experiment in other contexts during the past two decades. Our current understanding—at least of the right questions if not all the answers about superconductivity—would not have been possible without the advances in these apparently unrelated fields during the period (roughly 1970–85) when research on superconductivity was relatively dormant.

Flux lines and vortices

When a superconductor is cooled below its transition temperature, the electrons pair together to form 'Cooper pairs'. Although the BCS microscopic theory explains why this pairing occurs, the older, simpler and more phenomenological Ginzburg-Landau theory is adequate and indeed necessary to understand many of the most striking features of the superconducting phase¹. It focuses exclusively on the coherent quantum-mechanical wavefunction, ψ , of the Cooper pairs, its variations in space and time, and its interactions with electric and magnetic fields. This wavefunction is a complex scalar and can thus be characterized by its amplitude and phase: $\psi = |\psi| e^{i\phi}$.

In general both the amplitude and the phase of the wave function will vary in space and time, but it is primarily the spatial correlations in the phase, ϕ , that underlie the striking features of superconductivity. In the absence of a magnetic field, as a material is cooled into the superconducting state, the phase

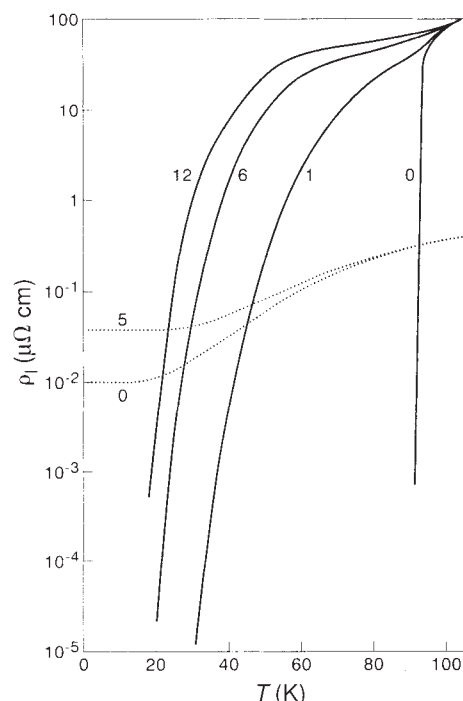


FIG. 1 The temperature (T) dependence of the linear response resistivity, $\rho_l \equiv \lim_{J \rightarrow 0} \{E/J\}$, where E is the electric field and J is the current density in the material. The solid curves are for $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (BSCCO) and, for comparison, the dotted curves are for high-conductivity copper. Each curve is for a different magnetic field, indicated in tesla. Data are from refs 33–35. For low-temperature superconductors the drop in the resistivity remains fairly abrupt even in a magnetic field (as it is for BSCCO in zero field only).

of the wavefunction becomes macroscopically correlated, taking on the same time-averaged value at every point in the material. Such macroscopic correlations are referred to as 'long-range order', a concept which plays a central role in modern condensed-matter physics.

Another classic example of long-range order occurs in magnetic materials. These come in a myriad of varieties, ranging from ferromagnets and antiferromagnets to the exotic frustrated materials called spin glasses². There are fruitful analogies between the types of magnetic order and the ordered states of a superconductor when it is placed in a magnetic field, as indicated in Fig. 2. The spins of the electrons in a solid can be visualized as tiny microscopic magnets. In a ferromagnet, the interactions between spins cause them to align in the same direction throughout the material. This cooperative phenomenon produces a permanent magnet, with each spin on average pointing towards, say, the north pole of the magnet. Such magnetic long-range order is closely analogous to the long-range order in a superconductor, with the direction of ψ in the complex plane (the phase ϕ) playing the role of the spin's orientation.

One of the classic signatures of a superconductor, and one that is often used in testing new materials for possible superconductivity, is the expulsion of a small external magnetic field—the Meissner effect (Fig. 2). Although superconductors expel small magnetic fields, a sufficiently intense field will penetrate the material, but not without unusual consequences. In type I superconductors, the magnetic field can only penetrate at the cost of destroying the superconductivity. In type II superconductors, there is a compromise: small fields are expelled, but a magnetic field in excess of the lower critical field, H_{c1} , penetrates nonuniformly, forming the 'mixed state'. The magnetic field forms flux lines, each line carrying precisely one unit of magnetic flux, the flux quantum $\phi_0 = hc/2e$ which is set by the total charge $2e$ of the Cooper pairs. The diameter of each flux line is set by a characteristic length, called the magnetic penetration length, and denoted λ . More generally, λ is the length scale over which a magnetic field can vary in a superconductor. Near the middle of each flux line the amplitude, $|\psi|$, of the Cooper-pair wavefunction is suppressed to zero. The region where the amplitude is strongly suppressed is called the vortex core and has a radius of size the coherence length ξ , the second fundamental length characterizing a superconductor. In the high- T_c copper oxides, ξ , which is in the range 10–20 Å at low temperatures, is much smaller than the penetration length, λ , which exceeds 1,000 Å (ref. 3). In low- T_c type II superconductors, ξ is usually much larger. The ratio $\kappa \equiv \lambda/\xi$ is an important dimensionless parameter. In Ginzburg–Landau theory, the value $\kappa = 1/\sqrt{2}$ marks the boundary between type I and type II materials. The high- T_c cuprates, with very large κ , are extreme type II superconductors. Here we focus only on type II superconductors, which enter the mixed state when a magnetic field penetrates the material.

On fully encircling a vortex outside its core, the phase ϕ of the wavefunction changes (winds) by 2π . This phase change reflects electric currents (supercurrents) that flow around the vortex, screening the magnetic field and confining it to within a distance λ of the vortex core. In contrast to vortices in classical fluids, the vorticity in a superconducting vortex is quantized; the phase change around a vortex must be an integer multiple of 2π because the wavefunction ψ is single-valued. Because of this quantization, a superconducting vortex is topologically stable, and a vortex line cannot terminate in the interior of a superconductor.

Abrikosov⁴ elucidated the nature of the vortex or flux lines and predicted that when a type II superconductor is placed in a magnetic field in excess of H_{c1} , the vortices that penetrate the material should form a regular lattice. This prediction was confirmed a decade later⁵ in magnetic decoration experiments that showed a triangular array of flux lines, an Abrikosov vortex lattice (Fig. 2). If one neglects thermal fluctuations⁶, the

Abrikosov vortex lattice shows two kinds of long-range order, although as we shall see, neither is robust enough to survive the ever-present imperfections and fluctuations in real materials. The most apparent are the long-range translational (and orientational) correlations present in the vortex array itself, directly analogous to crystalline correlations in a solid. A more subtle long-range order is also present: the phase, ϕ , of the wavefunction is ordered in a nontrivial spatial pattern, reflecting the underlying vortex lattice. This long-range phase coherence is analogous to antiferromagnetic order in magnetic materials, in which the electron spins show order with a periodic spatial structure (see Fig. 2).

If the magnetic field in the vortex lattice is increased, the vortices become more closely spaced and their cores start to overlap. At the upper critical field the vortex lattice and the pairing of electrons disappear, and the material becomes 'normal'. Neglecting thermal fluctuations, the upper critical field is $H_{c2} = \phi_0/(2\pi\xi^2)$, so small coherence lengths give rise to large upper critical fields. For the high- T_c cuprate superconductors at low temperatures, the small coherence lengths should yield upper critical fields exceeding 100 T, greater than the fields available in today's magnets. A full phase diagram of a clean type II superconductor, obtained from a 'mean-field' treatment of Ginzburg–Landau theory, is shown in Fig. 3a for a particular value of κ . Similar phase diagrams are quantitatively good for most clean low- T_c type II superconductors. But thermal fluctuations and sample imperfections, which were omitted in deriving Fig. 3a, can greatly change this phase diagram⁷.

Thermal fluctuations

In low- T_c superconductors, thermal fluctuations generally have a minor role, but thermal fluctuations are much more important

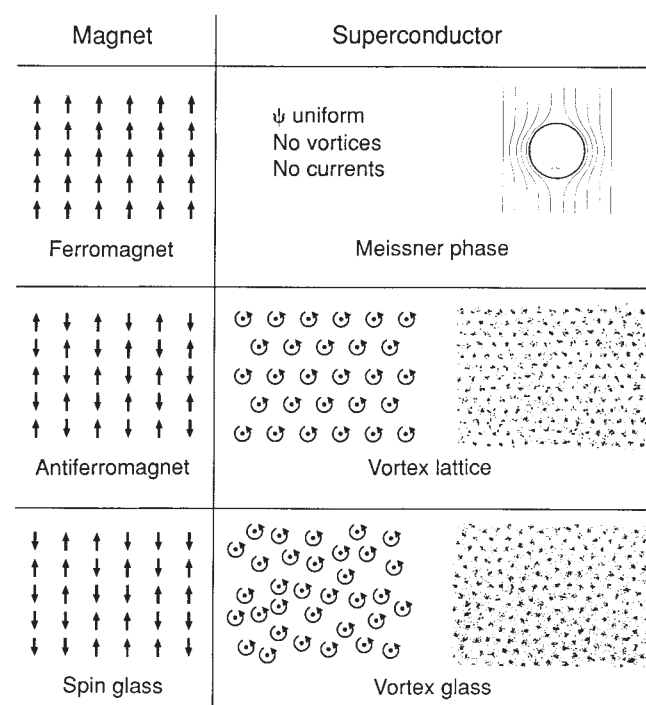


FIG. 2 Patterns of spin order in magnetically ordered systems (left, where arrows indicate the orientation of the local spins) and the analogous ordered phases of a superconductor (right). In the Meissner phase of the superconductor (top) the magnetic field and the vortices are expelled from the material, as indicated by the magnetic field lines around an ideal spherical superconducting ball. For the vortex lattice and glass, the positions of the vortex cores are indicated by dots and the circulating currents by arrows in the schematic drawings; on the right are images of vortex patterns in BSCCO crystals obtained by decorating the surface with fine magnetic particles that preferentially stick near the magnetic flux lines on each vortex. The decoration images are from the work reported in ref. 36.

in the high- T_c copper oxide superconductors. This is because the temperatures are higher and the energies to create and move vortices are lower. The relevant energy scale is that required to create a piece of vortex line whose length is one coherence length⁷. This energy scale varies as ξ/λ^2 , so it is much lower in the high- T_c materials with their small coherence lengths and large penetration lengths. The coherence length, ξ , is roughly the size of the bound Cooper pairs of electrons, which are much smaller in the high- T_c materials, and the large penetration lengths reflect weak supercurrents because of the relatively low density of mobile electron pairs.

The most striking effects of enhanced thermal fluctuations in the high- T_c superconductors are found in an applied magnetic field. The phase diagram for a type II superconductor with strong thermal fluctuations is shown in Fig. 3b. Notice the vortex-fluid regime between the mean-field upper critical field, $H_{c2}(T)$, and the vortex-glass phase. On cooling in a field, the electrons start to pair and vortices form in the pair wavefunction near H_{c2} , but the vortices do not freeze until substantially lower temperatures. The existence of a substantial vortex-fluid regime had been noted for thin superconducting films^{8,9}, but it was not observed in bulk superconductors until the copper oxide superconductors had been studied^{10,11}. In the vortex fluid, vortices are mobile and have only short-range correlations in their positions, much like atoms in a conventional fluid. The vortex-fluid regime extends to particularly low temperatures in extremely anisotropic layered materials (like BSCCO) when the magnetic field is directed perpendicular to the layers^{12,13}. In this case, the vortex lines actually consist of strings of point or 'pancake' vortices in each superconducting layer, with only rather weak correlations between vortices in different layers. The presence of a large vortex-fluid regime explains quite naturally the resistivity data for BSCCO shown in Fig. 1.

The vortex fluid is the natural analogue of the disordered, paramagnetic phase that occurs in magnetic materials heated above the Curie (or Neel) temperature. As in a paramagnet, the vortex fluid does not have any long-range order: vortex motion 'scrambles' the phase ϕ of the pair wavefunction, disrupting any possible long-range phase coherence and destroying superconductivity. It should be emphasized that the vortex fluid is not really a distinct phase: its properties evolve smoothly and continuously as the field or temperature is increased through H_{c2} to the normal state.

Resistivity

Which of the phases in the phase diagrams (Fig. 3) are actually superconducting, with zero resistance? Although zero resistance is in many ways the defining characteristic of a superconductor, it is in some sense the hardest to explain. A crucial part is played by the motion of vortex lines, which causes dissipation and thus resistivity. In a superconductor the local electrical current flows in the direction in which the phase ϕ increases and the current density $\mathbf{J}$ is proportional to the rate of increase of ϕ along that direction: $\mathbf{J} \propto \nabla \phi$. This current exerts a Magnus force on a vortex line, pushing it across the current (a closely related effect gives an aerofoil upward lift when a wind blows across it). The total change of phase ϕ on passing through a superconducting material on one side of a vortex line differs by 2π from that found on passing on the other side. When vortex lines move across the material in response to the Magnus force they move in the direction that reduces the total change of phase and thus reduces (dissipates) the current. To maintain a steady current, a voltage difference (and thus an electric field E) must be maintained across the material. The voltage difference acts to increase the phase difference across the material, balancing the decrease due to the motion of the vortex lines, and thus maintaining a steady current. In the Meissner phase ($H < H_{c1}$) the magnetic field is expelled from the material and there are no free vortex lines present to move and cause resistivity. Thus the Meissner phase is indeed superconducting with zero linear resis-

tivity $\rho_l \equiv \lim_{J \rightarrow 0} \{E/J\} = 0$. Note that ρ_l is the ohmic resistivity which measures the linear response to an applied current or field. As we shall see later, even in the Meissner phase there is some nonlinear resistivity: a nonzero electric field is needed to maintain any nonzero current density in the interior of the material, so the nonlinear resistivity $\rho \equiv E/J$ only vanishes in the limit of zero current density.

In a superconductor in the mixed state ($H > H_{c1}$), there are vortex lines present, induced by the penetrating magnetic field. Motion of these lines leads to resistivity. In the vortex-fluid regime the vortex lines are mobile because of thermal fluctuations even without a current, and can thus move in response to a current, leading to a nonzero resistivity. But what about in the Abrikosov vortex-lattice phase? Perhaps surprisingly, in a perfectly clean and ideal material the entire vortex lattice would be free to move in response to a current, and again cause nonzero resistivity. In an ideal superconductor, therefore, the Abrikosov vortex-lattice phase is not really superconducting. Nevertheless, real materials always have chemical and structural imperfections ('dirt') and these can impede the motion of the vortex-lattice¹⁴. In the presence of a penetrating magnetic field, if we wish to answer the question posed in our title, we must consider the role played by dirt.

Dirt and the vortex-glass phase

More than 20 years ago, Larkin¹⁵ argued that material imperfections destroy the crystalline long-range order of the Abrikosov vortex lattice. There are two competing effects: the interactions between the vortex lines favour a lattice structure, but the

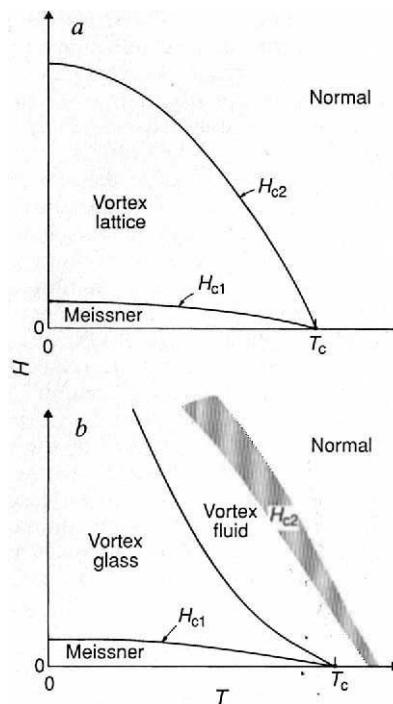


FIG. 3 a, Phase diagram of a type II superconductor, ignoring effects of thermal fluctuations and material imperfections, as a function of temperature, T , and applied magnetic field, H . This phase diagram is qualitatively correct for most low- T_c type II superconductors. b, Schematic phase diagram for a type II superconductor with strong thermal fluctuations, such as the high- T_c cuprates. In the absence of material imperfections the low-temperature ordered phase in a magnetic field above H_{c1} would have been a vortex lattice but the ever-present imperfections destabilize the vortex lattice, replacing it with the vortex glass phase, as shown. The transition at H_{c2} in a has here been broadened by fluctuations and has become only a gradual crossover from a normal metal to a vortex fluid, indicated by the shaded region.

randomly located material defects tend to pin the lines at random positions, disrupting the lattice structure. Larkin argued that beyond a characteristic length scale, l_L , the random vortex pinning dominates and destroys the vortex-lattice phase. Although the size, l_L , of the microcrystalline regions can be very long for weak pinning, the long-range lattice (positional) correlations are destroyed for distances in excess of l_L .

The conventional theory of vortex motion, based on ideas originated in the early 1960s by Anderson and Kim^{14,16,17}, maintains that finite 'bundles' of vortex line segments of length and diameter of order l_L can move by thermal activation across the 'landscape' of free energy barriers caused by the pinning. In this approach the interactions between the vortex bundles are ignored. Because the bundles are finite, the barriers have a finite maximum height, U . The independent motion of vortex bundles forced by an applied current would destroy the phase coherence of the pair wavefunction and lead to a finite resistivity with a thermally activated (Arrhenius) form: $\rho_1 \sim \exp(-U/k_B T)$. Thus the macroscopic behaviour of vortices in the mixed state was predicted to be that of a very sluggish vortex fluid (like cold molasses flowing through a sponge). According to this theory, superconductors in a magnetic field are not really superconducting, except at zero temperature. For practical purposes, however, this conclusion is almost irrelevant for conventional superconductors: $U/k_B T$ is so large that ρ_1 is immeasurably small except extremely close to the H_{c2} line.

For the high- T_c copper oxide superconductors, on the other hand, $U/k_B T$ is not large and one must reconsider the issue, particularly in light of the new experimental measurements (for example those in Fig. 1). We have recently explored a different scenario^{7,18} from that envisaged by Anderson and Kim. We argue that on cooling, the fluid of vortex lines interacting with the random pinning sites undergoes a sharp thermodynamic phase transition into a truly superconducting phase with no mobile vortices and thus strictly zero resistivity. In this superconducting phase the vortices are frozen into a sample-specific compromise configuration, determined in detail by the competition between the interactions of the vortices with each other and with the microscopic impurities in the material. Because the vortices are immobile, the phase of the pair wavefunction has true long-range order, although in a seemingly random pattern which reflects in detail the specific positions of the frozen vortices. The order present in this superconducting phase is directly analogous to the magnetic order that occurs in some random magnetic alloys, called spin glasses (see Fig. 2). In spin glasses², the spin axes of the electrons are frozen in time, but rather than being aligned as in a ferromagnet, or in a spatially periodic pattern as in an antiferromagnet, they are oriented in a sample-specific arrangement, determined by the microscopic details of the material. Because of this analogy, the new superconducting phase has been named the 'vortex glass'¹⁹.

In contrast to the Meissner phase, the superconducting vortex-glass phase can only exist in an impure material. Thus we argue

that superconductors in a penetrating magnetic field can be truly superconducting only when they are dirty. The random pinning which is present to some extent in any real material, no matter how carefully prepared, destabilizes the vortex-lattice phase (which was not truly superconducting) and converts it into the superconducting vortex-glass phase.

In our view, the fundamental flaw in the theories of thermally activated vortex motion based on Anderson and Kim's work^{14,16,17} is that they assume independent motion of finite bundles of vortex-line segments. But as mentioned above, a vortex line cannot end in a superconductor, so the vortex lines are constrained to remain connected as segments of them are moved. We argue that at low temperatures this constraint and the interactions between the vortex lines (which extend out to distances of the order of the penetration length, λ) cannot be ignored, even on length scales well beyond the microcrystallite size l_L . These interactions cause phase correlations which at sufficiently low temperature can propagate to arbitrarily long distances, producing the long-range order of the vortex-glass phase.

We will now consider some of the key predictions of the theory of the vortex-glass phase and their experimental confirmation. The evidence for the vortex-glass phase comes from transport experiments on the copper oxide superconductors conducted over the past few years²⁰⁻²⁴.

Vortex-glass phase transition

A fruitful way to distinguish a truly superconducting phase, with strictly zero resistivity, from a vortex fluid with immeasurably small but nonzero resistivity is to search for a phase transition as the temperature is lowered. If the superconducting vortex-glass phase exists, there must be a sharp thermodynamic phase transition which separates it from the (nonsuperconducting) vortex fluid. This phase transition may be of first order, in which case the resistivity will jump discontinuously to zero and other properties will be discontinuous at the phase transition, or it may be continuous (second order). In the latter case, which seems to obtain for samples of $\text{YBa}_2\text{Cu}_3\text{O}_7$ (YBCO) with strong enough random pinning, one expects universal critical scaling behaviour for temperatures near the vortex-glass phase transition temperature, T_{vg} .

Critical scaling behaviour occurs in all continuous phase transitions²⁵ and can be attributed to the existence of a characteristic length which diverges at the transition. Above the vortex-glass transition, the relevant length, ξ_{vg} , is the length scale over which the phase of the pair wavefunction is correlated (albeit in a random pattern). As at most continuous phase transitions, one expects a power-law divergence of the correlation length as the transition temperature is approached: with $\xi_{vg} \sim |T - T_{vg}|^{-\nu}$, characterized by a critical exponent ν . Critical slowing of the dynamics is also expected near a continuous phase transition, with the relaxation time, τ , diverging as a power of ξ_{vg} , $\tau \sim \xi_{vg}^z \sim |T - T_{vg}|^{-z\nu}$, where z is the dynamical critical

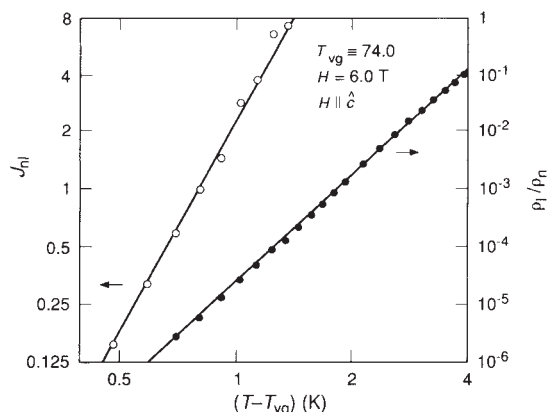


FIG. 4 Scaling in the vortex-glass critical regime for a crystal of YBCO in a 6-T magnetic field applied parallel to the material's $\hat{c}$ -axis (perpendicular to the CuO_2 planes). Data are from ref. 21. The solid circles are the linear resistivity, ρ_1 , divided by the resistivity in the normal state, ρ_n , plotted on logarithmic scales against the difference of the temperature, T , from that of the vortex-glass phase transition, $T_{vg} = 74$ K. The slope of the solid line through the solid circles yields the critical exponent $s = \nu(z - 1) \approx 6.5$. The open circles are the current density, J_n (in arbitrary units), for the onset of nonlinearity in the resistivity. Here the slope yields $2\nu \approx 4$.

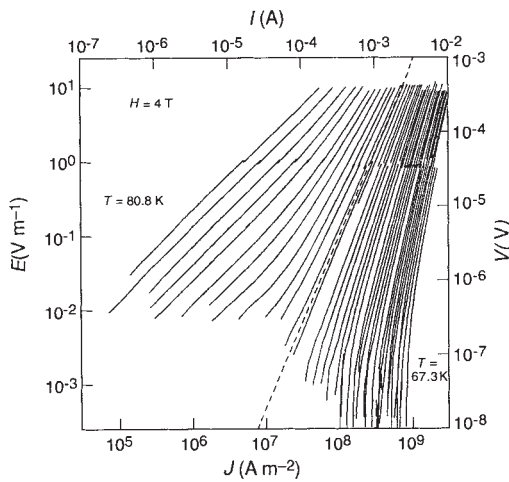


FIG. 5 Current-voltage (I - V) curves at fixed temperatures for a 0.4- μm -thick YBCO epitaxial film in a 4-T magnetic field applied parallel to the material's c -axis. Adjacent curves differ by temperature intervals of ~ 0.3 K, with the highest and lowest temperatures also indicated. The curves are on logarithmic scales, and the current densities J and electric fields E are also indicated. The dashed line indicates the vortex-glass transition temperature at $T_{vg} \approx 77$ K at which $E \sim J^{(z+1)/2}$ with $z \approx 4.8 \pm 0.2$. For temperatures above T_{vg} at low current densities, $E \sim J$, indicating the ohmic behaviour characteristic of a normal metal or vortex fluid. Below T_{vg} , the slope of the curves becomes steeper and steeper at low voltages suggesting exponentially small dissipation as predicted for the vortex-glass phase. Data are from ref. 20. The curves were measured in separate sweeps up to $E \approx 1 \text{ V m}^{-1}$ and to $E \approx 10 \text{ V m}^{-1}$. The breaks between the two sets of curves at $E \approx 1 \text{ V m}^{-1}$ are due to resistive heating in the larger E sweeps.

exponent. This time is, roughly speaking, the time it takes a fluctuation of size ξ_{vg} to relax. The renormalization group theory of critical phenomena near continuous phase transitions²⁵ predicts that the critical exponents are universal numbers which characterize the type of transition but are insensitive to microscopic details, such as the material. In addition, one expects there to be universal scaling laws which can relate measurements at different temperatures near the transition, provided each physical quantity is scaled by the appropriate power of $(T - T_{vg})$. In particular we will focus on the dissipation as a function of applied current: the current-voltage curves. The linear resistivity is predicted to vanish as $\rho_l \sim (T - T_{vg})^s$ as the transition is approached from above, with $s = \nu(z - 1)$, an example of a scaling law⁷. Such a power-law form with exponent $s \approx 6.5$ for the resistivity has been observed for a sample of YBCO over a temperature range where the resistivity drops by more than four orders of magnitude²¹ (Fig. 4).

At temperatures just above T_{vg} , the vortex-fluid state has a very long correlation length ξ_{vg} . But these correlations can be disrupted by the vortex-line motion caused by even a small applied current density J , so the nonlinear resistivity $\rho(J) \equiv E/J$ is very sensitive to the magnitude of J at temperatures near T_{vg} . It is useful to define a characteristic current-density scale J_{nl} as that current at which the vortex-fluid state is significantly disrupted. To be specific, for $T > T_{vg}$ we choose J_{nl} as the current at which the nonlinear resistivity doubles: $\rho(J_{nl}) = 2\rho_l$. We predict⁷ that J_{nl} should vanish as the inverse square of the correlation length, $J_{nl} \sim \xi_{vg}^{-2} \sim (T - T_{vg})^{2\nu}$; the measurements of Gammel *et al.*²¹ on YBCO yield $2\nu \approx 4$ (Fig. 4). This rapid drop of J_{nl} as T approaches T_{vg} from above is the experimental result most strikingly inconsistent with the 'conventional' theories^{14,16,17} of thermally activated vortex motion. These theories assume that the correlation length is roughly Larkin's crossover length l_L , so they do not predict any strong temperature dependence of J_{nl} in this regime. The 'conventional' approach therefore fails at temperatures and current densities low enough that the corre-

lation length, ξ_{vg} , is greater than the size of the microcrystalline regions, l_L .

In addition to predicting these power-law behaviours, scaling theory predicts that the functional dependence of the electric field on the current density will be the same for any temperature near but above T_{vg} , provided that the dissipation is normalized by the linear resistivity and the current density is normalized by J_{nl}

$$\frac{\rho(J)}{\rho_l} \equiv \frac{E}{J\rho_l} \approx \mathcal{R}_+ \left(\frac{J}{J_{nl}} \right) \quad (1)$$

where $\mathcal{R}_+(j)$ is a universal scaling function with $\mathcal{R}_+(j \rightarrow 0) = 1$ and, from the above definition of J_{nl} , $\mathcal{R}_+(1) = 2$. As we decrease T to T_{vg} at fixed J , both $\rho(J)/\rho_l$ and J/J_{nl} diverge. For E and $\rho(J)$ at T_{vg} to be well-behaved, the scaling function in equation (1) must behave as $\mathcal{R}_+(j) \sim j^{(z-1)/2}$ for large j . Then at T_{vg} one finds $\rho(J) \sim J^{(z-1)/2}$ or $E \sim J^{(z+1)/2}$, a power-law relationship between voltage and current. Current-voltage (I - V) curves measured by Koch *et al.*²⁰ on YBCO show the general features of the continuous phase transition (Fig. 5). The power-law behaviour at T_{vg} is indicated by the dashed line. For $T > T_{vg}$ and $J \ll J_{nl}$, ohmic behaviour with $E \approx \rho_l J$ is observed, whereas for larger currents the nonlinearity becomes appreciable. Both ρ_l and J_{nl} decrease continuously as T approaches T_{vg} from above. The values of the critical exponents extracted from the two experiments^{20,21} described above are in good agreement, although the samples and current densities studied were different. Moreover, recent computer simulations^{26,27} of a model superconductor reveal evidence for a vortex-glass phase transition with similar critical exponents.

Non-ohmic resistance

Just below the vortex-glass transition at T_{vg} there is a scaling behaviour⁷ analogous to equation (1), with a characteristic current scale J_{nl} . But now the behaviour for $J \ll J_{nl}$ is not ohmic; rather, as the current is decreased the voltage vanishes faster than any power of the current, as indicated by the curvature to very high slopes at the bottom right of Fig. 5. This dissipation regime is that due to vortex motion in the ordered vortex-glass phase. To understand this, let us first consider the analogous dissipation processes in the much simpler Meissner phase.

In the Meissner phase ($H < H_{c1}$) there are no free vortices present, but at nonzero temperatures thermal fluctuations spontaneously occur, typically making local excitations with free energy of the order of the thermal energy, $k_B T$. One type of excitation that occurs and can lead to dissipation of a current

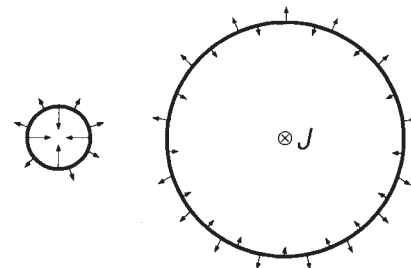


FIG. 6 The bold circles represent closed vortex loop excitations in the Meissner phase. The inward-pointing arrows denote the inward force on the loop due to the line tension of the vortex; this force is proportional to the curvature of the loop. In the presence of a current density passing through the loop (into the page) there is also an outward Magnus force on the loop due to its interaction with this current (the opposite current would produce an inward force). The loop on the left is smaller than the critical radius so the inward force is larger and the loop tends to contract. The loop on the right is larger than the critical radius so the outward force dominates, and it tends to grow still larger, contributing to the dissipation of the current.

is closed vortex loops of finite circumference²⁸ (Fig. 6). The free energy, $f_v(R)$, of a circular vortex loop of radius R is roughly proportional to its circumference, $f_v(R) \approx 2\pi R\epsilon$, where ϵ is the free energy per unit length or, equivalently, the line tension of a vortex. The Boltzmann probability of this loop appearing as a spontaneous thermal fluctuation at temperature T is thus $\exp[-f_v(R)/k_B T]$. This is the probability that random thermal fluctuations coincidentally produce outward forces on the loop of sufficient average strength and duration for it to grow to this size. In the absence of any applied current such a vortex loop will almost certainly rapidly shrink in response to the ever-present inward force of ϵ/R per unit length arising from its line tension and curvature. Let us assume, however, that there is a uniform current density J passing through the loop. The outward Magnus force from this current can counterbalance the inward force from the vortex line tension (Fig. 6) and cause loops of radius greater than a critical size $R_c(J) \approx \epsilon/J$ to tend to grow indefinitely; the resulting vortex motion produces dissipation²⁸. (The process of nucleation of vortex loops bigger than the critical radius $R_c(J)$, is analogous to bulk nucleation of droplets of liquid in a supersaturated vapour, the applied current playing the role of the degree of supersaturation and R_c the role of the critical droplet radius.) The dissipation rate in the Meissner phase is proportional to the probability of nucleating sufficiently large loops, $\exp[-f_v(R_c(J))/2k_B T]$, yielding, for small current density J

$$E/J = \rho(J) \approx \exp[-(J_T/J)^\mu] \quad (2)$$

with $\mu = 1$ and $J_T \approx \epsilon^2/k_B T$ setting the current scale of this thermally activated dissipation. Thus there is always dissipation, even in the Meissner phase, but at small current densities it arises from rare large thermal fluctuations and vanishes exponentially for $J \rightarrow 0$.

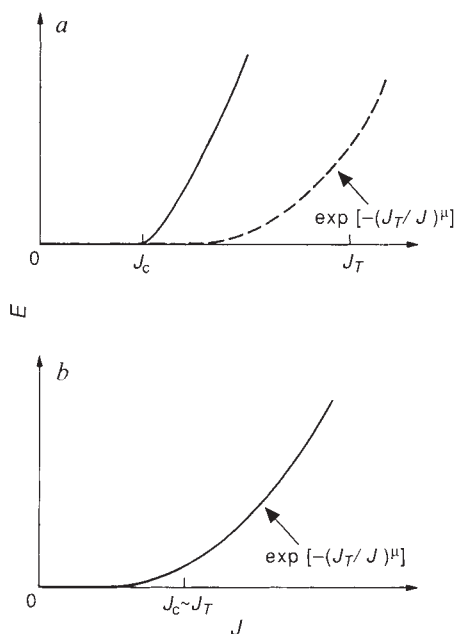


FIG. 7 Electric field E against current density J in a truly superconducting phase (a Meissner or vortex-glass phase). *a*, With weak thermal fluctuations, $J_T \gg J_c$ and the electric field for $J < J_c$ is small. The dashed line is the extrapolation of the activated behaviour for $J < J_c$ to $J > J_c$. Near J_c , however, the activation barriers for vortex motion vanish and there is a rapid increase in the dissipation, leading to an apparently sharp critical current. This behaviour is found in low- T_c superconductors. *b*, With strong thermal fluctuations, J_c and J_T are comparable and the onset of dissipation is more gradual. This behaviour is found in the high- T_c superconductors (except at temperatures far below T_c where the behaviour is more like that in *a*).

How does this fit in with the conventional concept of a 'critical current' in superconductors? For small current densities, the free energy barrier that must be crossed to produce a vortex loop varies as $J^{-\mu}$. But this barrier becomes of order $k_B T$ at the critical current J_c and the thermally activated dissipation law of equation (2) thus only applies for $J < J_c$; for $J > J_c$ vortices are readily produced, yielding strong dissipation. In low- T_c superconductors, $J_c \ll J_T$ except very near T_c , so the nucleation of loops is very rare for $J < J_c$. Thus one finds a fairly abrupt increase in the dissipation at J_c (Fig. 7*a*). Very near T_c in conventional superconductors, and over large portions of the phase diagram of the high- T_c copper oxides, on the other hand, J_c and J_T are comparable so the onset of dissipation is more gradual, with the critical current ill defined (Fig. 7*b*).

In the vortex-glass phase, in which many vortices penetrate the sample, qualitatively similar dissipative behaviour occurs, although the details are much more subtle and not yet fully understood. It is not clear what the precise character of the dominant vortex-line rearrangements leading to nonlinear dissipation will be in the vortex-glass phase; several proposals have been analysed^{7,29-31}. In each of the proposed dissipative processes, however, the free energy barriers that must be crossed to produce the vortex rearrangements grow as a power of the size of the rearranged region, just as in the Meissner phase. If a current is applied through the superconductor, it will induce the thermally activated production of such rearrangements larger than a critical size $R_c(J)$. This will cause phase slip and hence dissipation as in the Meissner phase leading again to a low-current resistance of the form $E/J \approx \exp[-(J_T/J)^\mu]$, but now with an exponent $0 < \mu \leq 1$ determined by the nature of the dissipative processes. Such thermally activated vortex motion is often called vortex (or flux) creep. As in the Meissner phase, a sufficiently strong current can overcome the free energy barriers to vortex motion, causing vortex (or flux) flow and rapid dissipation: this will occur above the 'critical' current density J_c . We thus conclude that both the Meissner and vortex-glass phases are linear superconductors with vanishing resistivity only in the linear-response limit of zero current density. For any nonzero current density, the resistivity is nonzero because of thermally activated vortex motion, although it vanishes as an exponential function of the current density, J , for $J \rightarrow 0$.

Because of the very small voltages that occur in the activated regime, direct measurement of the behaviour in equation (2) and experimental determination of the exponent μ is difficult. Recent measurements^{20,23,24} in the vortex-glass phase, however, yield reasonable fits to equation (2) with values of μ less than $1/2$.

Another way to study nonlinear dissipation in the vortex-glass phase is through magnetization relaxation at low temperatures. In a typical experiment a sample is cooled into the vortex-glass phase in a magnetic field, and then the field strength is suddenly changed. This change induces a non-equilibrium screening current, $J(t)$, which decays to zero as vortex lines enter or leave the material and move to their new equilibrium configuration. The slow decay of this current, which can be measured through the associated magnetization, is given by $dJ(t)/dt \propto E(J)$, with $E(J)$ the electric field needed to sustain the current J in steady state. In conventional superconductors (and copper oxide superconductors at very low temperatures) dJ/dt is so small for $J < J_c$ that the current density never drops much below J_c , even in a year. With the strong fluctuations in the copper oxide superconductors, by contrast, the current can decay much further; and the predicted form for $E(J)$ in equation (2) yields a decay as $J(t) \approx J_T [\ln(t/t_0)]^{-1/\mu}$ at very long times, with t_0 a microscopic time of order $10^{-9} - 10^{-12}$ s. This technique gave Sandvold and Rossel²³ a voltage sensitivity almost 10 orders of magnitude better than conventional transport measurements, corresponding to electric fields down to about 10^{-14} V cm⁻¹. Their inferred $E(J)$ is shown in Fig. 8. They find good fits to equation (2) in YBCO films with $\mu \approx 1/3$. This is consistent with the transport

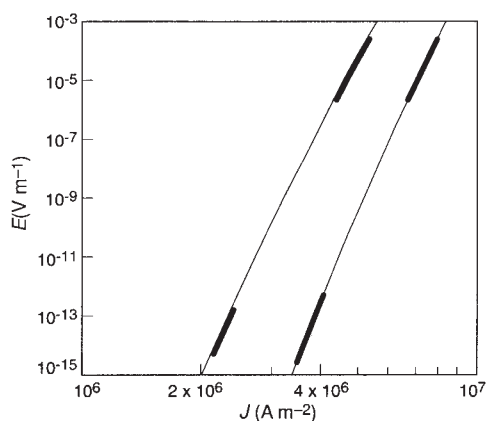


FIG. 8 Electric field E against current density, J on logarithmic scales, as measured for a 0.2- μm -thick epitaxial film of YBCO grown on $\text{SrTiO}_3(100)^{23}$. The applied magnetic fields are 0.3 T (left) and 0.1 T (right) and the temperature is 70 K, which is in the vortex-glass phase. The data are the thick curves; at high E the electric field was directly measured, whereas at very low E it was inferred from the slow decay with time of circulating currents in the material. The fits to equation (2) with $\mu=0.34$ are indicated by the fine lines.

measurements and, together, these experiments provide strong evidence that the resistance vanishes faster than any power of the current in the vortex-glass phase.

Conclusions

We have seen that the combined effects of thermal fluctuations and impurities lead to phase diagrams and transport properties of the high- T_c copper oxide superconductors in a magnetic field that are radically different from those of low- T_c type II superconductors. If a conventional low- T_c type II superconductor is cooled in a magnetic field in excess of H_{c1} , the resistivity abruptly becomes immeasurably small as the $H_{c2}(T)$ line in Fig. 3a is crossed. In striking contrast, because of the strong thermal fluctuations in the copper oxide materials, there is no well defined H_{c2} line but only a gradual pairing of the electrons into Cooper pairs leading to the formation of vortex lines (Fig. 3b). In the resulting non-superconducting vortex-fluid regime, as the temperature is lowered the vortex motion is increasingly impeded by impurities so the resistivity decreases. This continues until the vortex-glass phase transition, $T_{vg}(H)$, where the linear ohmic resistivity finally vanishes. Near T_{vg} scaling behaviour of current-voltage curves is found, providing the best evidence for a true phase transition. In the vortex-glass phase the dissipation of a small current is dominated by the thermal activation of large collective vortex rearrangements which cause a resistance that vanishes exponentially with current at small currents. At very low temperatures, this thermally activated dissipation is small, and current-voltage curves show a fairly abrupt onset of dissipation at the critical current, with resistance too small to be directly measured at lower currents; this regime is similar to low- T_c superconductors.

Although experiments are qualitatively, and in some respects quantitatively, in accord with the theoretical picture presented here, there remain many open questions and new regimes to be explored. These include the dependence of T_{vg} and the current scales J_T and J_c on the material and its type and distribution of imperfections, one interesting limit being that of very few imperfections (the clean limit). In addition, the nonlinear dissipation in the mixed state has not been carefully examined in either highly anisotropic, quasi-two-dimensional materials such as BSCCO or in isolated, truly two-dimensional superconducting thin films; the latter are expected to be truly superconducting only for zero magnetic field or zero temperature, with no vortex-glass phase at nonzero temperature³². Another interesting avenue is the study of artificially structured superconductors, as produced by microfabrication techniques such as molecular-beam epitaxy, where the properties of the structure might be systematically and continuously varied, for example by introducing nonsuperconducting layers of different thicknesses.

Finally, we should directly answer the question posed in our title. Thanks to experiments on the high- T_c copper oxide superconductors and to new theoretical ideas, the answer that was thought to be “no, superconductors in the mixed state are not really superconducting; their resistance is just extremely small” is in fact, “yes, they are really superconducting at temperatures below the phase transition into the vortex-glass phase”. $\square$

David A. Huse is at AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA. Matthew P. A. Fisher is at the IBM T. J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA. Daniel S. Fisher is at the Physics Department, Harvard University, Cambridge, Massachusetts 02138, USA.

1. *Superconductivity*, Vols 1 & 2 (ed. Parks, R. B.) (Dekker, New York, 1969).
2. Edwards, S. F. & Anderson, P. W. *J. Phys.* **F5**, 965-974 (1975).
3. Batlogg, B. in *Physics of High-Temperature Superconductors* (eds Maekawa, S. & Sato, M.) 228 (Springer, Berlin, 1992).
4. Abrikosov, A. A. *Zh. eksp. teor. Fiz.* **32**, 1442-1452 (1957); (Engl. transl.) *Sov. Phys. JETP* **5**, 1174-1182 (1957).
5. Trauble, H. & Essman, U. *J. appl. Phys.* **39**, 4052-4059 (1968).
6. Moore, M. A. *Phys. Rev.* **B45**, 7336-7345 (1992).
7. Fisher, D. S., Fisher, M. P. A. & Huse, D. A. *Phys. Rev.* **B43**, 130-159 (1991).
8. Huberman, B. A. & Doniach, S. *Phys. Rev. Lett.* **43**, 950-952 (1979).
9. Fisher, D. S. *Phys. Rev.* **B22**, 1190-1199 (1980).
10. Gammel, P. L., Schneemeyer, L. F., Waszczak, J. V. & Bishop, D. J. *Phys. Rev. Lett.* **61**, 1666-1669 (1988).
11. Nelson, D. R. & Seung, H. S. *Phys. Rev.* **B39**, 9153-9174 (1989).
12. Houghton, A., Pelcovits, R. A. & Sudbø, A. *Phys. Rev.* **B40**, 6763-6770 (1989).
13. Brandt, E. H. *Phys. Rev. Lett.* **63**, 1106-1109 (1989).
14. Anderson, P. W. & Kim, Y. B. *Rev. mod. Phys.* **36**, 39-43 (1964).
15. Larkin, A. I. *Zh. eksp. teor. Fiz.* **58**, 1466-1470 (1970); (Engl. transl.) *Sov. Phys. JETP* **31**, 784-786 (1970).
16. Kes, P. H., Aarts, J., van den Berg, J., van der Beek, C. J. & Mydosh, J. A. *Supercond. Sci. Technol.* **1**, 242-248 (1989).
17. Brandt, E. H. *Int. J. mod. Phys.* **B5**, 751-795 (1991).
18. Fisher, M. P. A. *Phys. Rev. Lett.* **62**, 1415-1418 (1989).
19. Shih, W. Y., Ebner, C. & Stroud, D. *Phys. Rev.* **B30**, 134-144 (1984).

20. Koch, R. H. *et al. Phys. Rev. Lett.* **63**, 1511-1514 (1989).
21. Gammel, P. L., Schneemeyer, L. F. & Bishop, D. J. *Phys. Rev. Lett.* **66**, 953-956 (1991).
22. Worthington, T. K. *et al. Phys. Rev.* **B43**, 10538-10543 (1991).
23. Sandvold, E. & Rossel, C. *Physica C* **190**, 309-315 (1992).
24. Dekker, C., Eidelloth, W. & Koch, R. H. *Phys. Rev. Lett.* **68**, 3347-3350 (1992).
25. E.g., Fisher, M. E. *Rev. mod. Phys.* **46**, 597-616 (1974).
26. Reger, J. D., Tokuyasu, T. A., Young, A. P. & Fisher, M. P. A. *Phys. Rev.* **B44**, 7147-7150 (1991).
27. Gingras, M. J. P. *Phys. Rev.* **B45**, 7547-7550 (1992).
28. Langer, J. S. & Fisher, M. E. *Phys. Rev. Lett.* **19**, 560-563 (1967).
29. Feigel'man, M. V., Geshkenbein, V. B., Larkin, A. I. & Vinokur, V. M. *Phys. Rev. Lett.* **63**, 2303-2306 (1989).
30. Fischer, K. H. & Nattermann, T. *Phys. Rev.* **B43**, 10372-10382 (1991).
31. Bouchaud, J.-P., Mézard, M. & Yedidia, J. S. *Phys. Rev. Lett.* **67**, 3840-3843 (1991).
32. Fisher, M. P. A., Tokuyasu, T. A. & Young, A. P. *Phys. Rev. Lett.* **66**, 2931-2934 (1991).
33. Wilson, M. N. *Superconducting Magnets*, 99 (Oxford Univ. Press, 1983).
34. Palstra, T. T. M., Batlogg, B., Schneemeyer, L. F. & Waszczak, J. V. *Phys. Rev. Lett.* **61**, 1662-1665 (1988).
35. Safar, H., Gammel, P. L., Bishop, D. J., Mitzi, D. B. & Kapitulnik, A. *Phys. Rev. Lett.* **68**, 2672-2675 (1992).
36. Grier, D. G. *et al. Phys. Rev. Lett.* **66**, 2270-2273 (1991).

ACKNOWLEDGEMENTS. D.S.F. is supported by the NSF through the Materials Theory Program and Harvard University's Materials Research Laboratory, as well as by the A. P. Sloan Foundation.

A small metalloribozyme with a two-step mechanism

Tao Pan & Olke C. Uhlenbeck

Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA

An RNA molecule consisting of an asymmetric internal loop of six nucleotides can be rapidly and specifically cleaved by Pb^{2+} in the presence of Mg^{2+} . The 5' cleavage product terminates with a 3' phosphomonoester generated from a 2',3'-cyclic phosphodiester reaction intermediate. This two-step reaction mechanism resembles that of many protein ribonucleases but has not previously been observed for reactions catalysed by RNA.

ALL ribozyme reactions either require or are greatly stimulated by divalent metal ions. In a few cases the metal ions promote proper folding of the RNA or directly participate in the chemical catalysis¹⁻³. The specific cleavage of yeast transfer RNA^{Phe} between U17 and G18 by Pb^{2+} has been studied extensively as a model for the role of metal ions in the chemical step of RNA catalysis⁴⁻⁷. For yeast tRNA^{Phe}, three tightly bound Pb^{2+} ions have been identified in the crystal structure^{4,6}, one of which is placed so that a Pb^{2+} -bound hydroxyl ion can remove the proton from the 2'-OH of ribose 17. Attack of the resulting oxyanion at the adjacent phosphorus atom cleaves the RNA chain, generating 2',3'-cyclic phosphate and 5'-hydroxyl termini.

We have developed an *in vitro* selection method to isolate RNA molecules that undergo autolytic cleavage with Pb^{2+} in the presence of Mg^{2+} (ref. 8). Many specific Pb^{2+} cleavage motifs of 30–50 nucleotides were found that cleaved with rates equal to or faster than the Pb^{2+} cleavage of yeast tRNA^{Phe}. Here we report that one of these cleavage motifs, consisting of an RNA duplex with an asymmetric internal loop of six nucleotides, not only catalyses the phosphodiester cleavage at a specific site, but also promotes the hydrolysis of the 2',3'-cyclic phosphate from the initial cleavage to produce specifically 3'-monophosphate. Catalysis of transesterification and hydrolysis reactions in a two-step mechanism by this Pb^{2+} ribozyme is common in protein ribonucleases, but seems to be unique among all ribozymes described so far. For instance, hammerhead, hairpin and hepatitis δ ribozymes seem to be unable to catalyse hydrolysis of the 2',3'-cyclic phosphate produced from the transesterification reaction⁹⁻¹¹.

Small, efficient Pb^{2+} cleavage motif

The cleavage of variant number 34 (Fig. 1c)⁸ by Pb^{2+} was of particular interest because the rate of cleavage was >10 times faster than that of yeast tRNA^{Phe} and cleavage seemed to occur in a double-stranded region unlike all previously characterized Pb^{2+} cleavage sites^{6,8}. To determine how much of this RNA molecule was required for Pb^{2+} cleavage, 5'-³²P-labelled RNA was partially hydrolysed at alkaline pH to produce a distribution of truncated molecules and subjected to cleavage with Pb^{2+} in the presence of Mg^{2+} . Analysis of the reaction products on a sequencing gel (Fig. 1a) revealed that the full-length RNA and molecules with 3' deletions through C49 had been cleaved by Pb^{2+} , but the alkaline hydrolysis product ending at C48 was not. This established C48 as the 3' boundary of the cleavage motif. A similar experiment with 3'-³²P-labelled RNA indicated that G20 was the 5' boundary of the cleavage motif (Fig. 1b). Because so much of variant number 34 can be removed without affecting Pb^{2+} cleavage (boxed residues in Fig. 1c), the tRNA-like secondary structure depicted in Fig. 1c may not be relevant for the cleavage reaction by Pb^{2+} . Inspection of the sequence of the required nucleotides suggests an alternative secondary structure that maintains the anticodon stem-loop, but forms a new helix and an asymmetric internal loop of six nucleotides (Fig. 1d). This structure places the Pb^{2+} cleavage site in a more likely single-stranded region and provides a rationale for the observed boundaries of the Pb^{2+} cleavage motif. When the new helix is truncated to less than three base pairs, it may not be sufficiently stable to permit the formation of the internal loop.

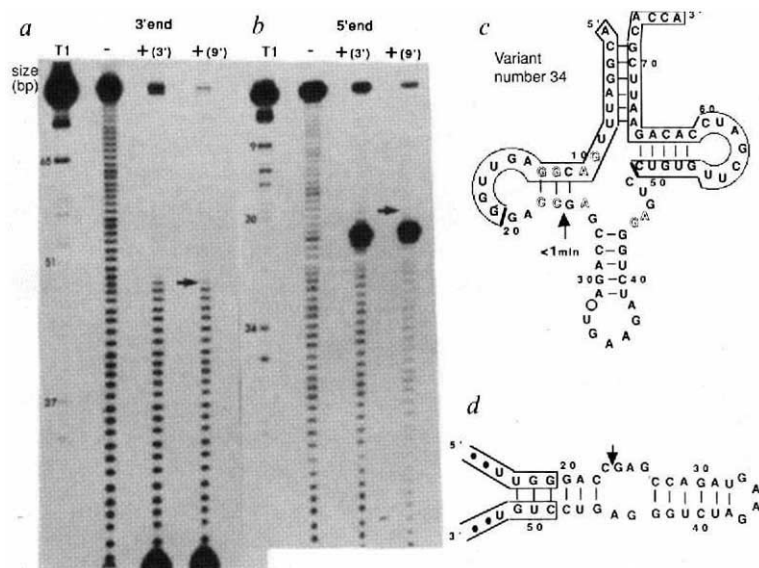
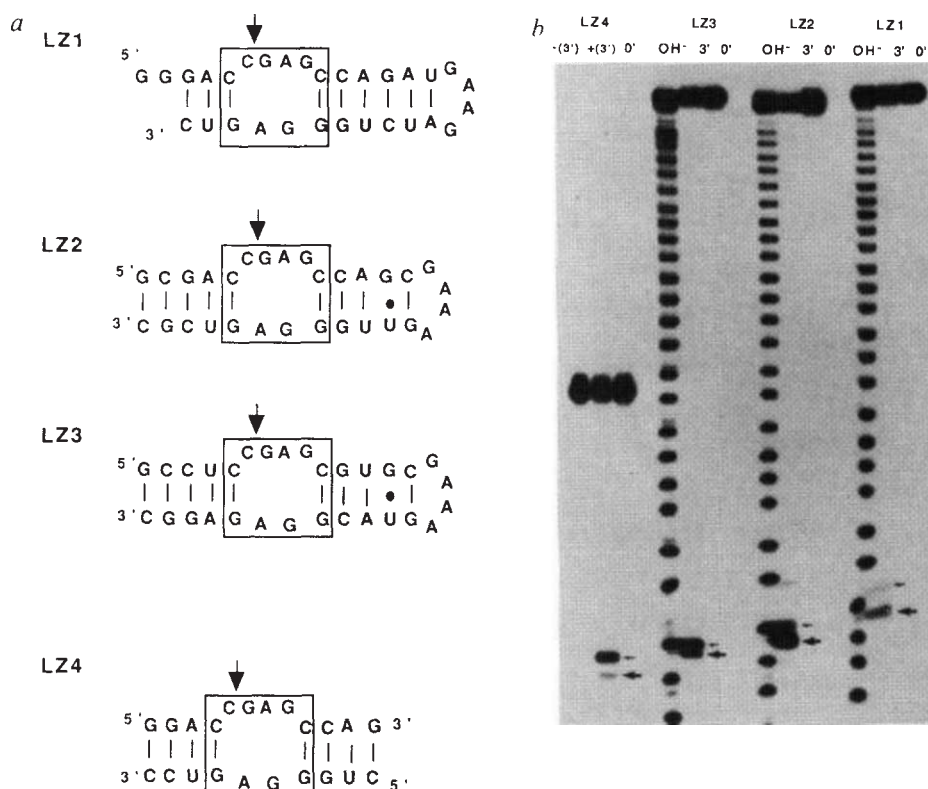


FIG. 1 Determination of the minimal Pb^{2+} cleavage motif of variant number 34 using the method as previously described^{8,18}. *a*, 3'-end boundary. Minus, No Pb^{2+} ; plus, Pb^{2+} was added and the reaction mixture was incubated at 25 °C for the time specified (in min) in parentheses; T1, partial ribonuclease T1 digestion of the same RNA. *b*, 5'-end boundary. *c*, Sequence of the variant number 34 (ref. 8) shown as cloverleaf secondary structure. The Pb^{2+} cleavage site is shown by an arrow and the time underneath corresponds to the apparent half-life of Pb^{2+} cleavage⁸. The nucleotide numbering is the same as that of yeast tRNA^{Phe} and the nucleotide changes are highlighted. A single deletion is shown as an open circle. Residues that can be deleted according to the experiments in *a* and *b* are boxed. *d*, Rearrangement of the secondary structure shown in *c*. The cleavage site is relocated in a single-stranded region of an asymmetric internal loop of six nucleotides.

FIG. 2 a, Sequences of RNA constructs that undergo Pb^{2+} cleavage at sites indicated by the arrows. The 10 residues identical in all four RNA constructs are boxed. b, Pb^{2+} cleavage of the RNA variants shown in a. In all cases, two $5'$ ^{32}P -labelled reaction products were obtained. The upper and lower bands are indicated by small and large arrows, respectively.

METHODS. LZ1-3 were obtained by *in vitro* transcription using T7 RNA polymerase¹⁹. The gel-purified RNAs were $5'$ ^{32}P -labelled using $[\gamma\text{-}^{32}P]\text{ATP}$ and T4 polynucleotide kinase. RNAs were renatured at $1\ \mu\text{M}$ in 15 mM MOPS, pH 7.0, by heating at 85°C for 2 min. MgCl_2 was added to 10 mM and cleavage initiated by addition of lead acetate to 0.2 mM (LZ1, 3) or 0.1 mM (LZ2). 0', No Pb^{2+} ; 3', incubation at 25°C for 3 min; OH⁻, partial alkaline hydrolysis ladder of the same $5'$ ^{32}P -labelled RNA. Both RNAs of LZ4 were chemically synthesized and gel-purified. The 12-mer at $1\ \mu\text{M}$ was incubated either alone (minus) or with 10 μM 10-mer RNA (plus) under conditions described above.



Although this alternative secondary structure may be in equilibrium with the tRNA-like structure shown in Fig. 1c, it is likely to be present in reasonably high proportions because several tertiary interactions that normally stabilize the D stem of tRNA are missing in variant number 34.

To verify the alternative secondary structure in Fig. 1d, four RNA constructs were prepared (Fig. 2a). LZ1 has the same sequence as variant number 34 and can form the minimal three base pairs of the proposed helix. In LZ2, both helices are designed to have five base pairs and the hairpin loop is changed to a stable GA_3 tetraloop. LZ3 differs from LZ2 by changing the sequences of two base pairs in each helix. LZ4 is similar to LZ1, but is formed by the combination of two separate RNAs and therefore does not contain the hairpin loop. All four RNAs cleaved rapidly with Pb^{2+} at the unique site (Fig. 2b). LZ1 cleaves at a significantly slower rate than the other three constructs, consistent with its less stable helix. In the case of LZ4, no cleavage of the 12-base oligonucleotide (12-mer) was observed unless the complementary 10-mer was present. These experiments clearly support the secondary structure in Fig. 1d and further define the sequence requirement for cleavage of the 10 boxed nucleotides in Fig. 2a that make up a purine-rich internal loop of six residues.

On high-resolution gels with $5'$ ^{32}P -labelled RNA, all four motifs show two product bands, an upper band that comigrates with and a lower band that migrates faster than the alkaline hydrolysis products (Fig. 2b). Because the RNA is $5'$ ^{32}P -labelled, it seems that the two products have distinct 3' termini. The construct LZ3 was most convenient for product identification because it could be prepared by *in vitro* transcription with $[\alpha\text{-}^{32}P]\text{GTP}$ and subsequently dephosphorylated to give rise to a $5'$ Pb^{2+} cleavage product which was uniquely ^{32}P -labelled at the 3' terminal phosphate. When the slow-migrating product was treated with nuclease P1 and alkaline phosphatase, analysis by thin-layer chromatography (Fig. 3a) revealed that 85–90% of the radioactivity comigrated with cytidine 2',3'-cyclic phosphate (CCP) with the remainder as inorganic phosphate. Thus, the upper product terminates with a 2',3'-cyclic phosphate residue, consistent with its comigration with the alkaline hydro-

lysis product. The small amount of inorganic phosphate is either the result of contamination by the other product or hydrolysis of cytidine 2',3'-cyclic phosphate to cytidine monophosphate by contaminating nucleases and subsequent dephosphorylation. When the faster-migrating Pb^{2+} cleavage product was digested with spleen phosphodiesterase II, 92–95% cytidine 3'-phosphate and 5–8% cytidine 2'-phosphate were obtained (Fig. 3b). Thus, the faster-migrating product terminates with a monophosphate with strong preference of the 3' over the 2' isomer.

2',3'-Cyclic phosphate reaction intermediate

Are the two 5' cleavage products with distinct 3' termini formed in separate reactions or is one an intermediate of the other? A time course of LZ1 cleavage showed that the amount of the product terminated with 2',3'-cyclic phosphate ($P2',3'\text{CP}$) increased rapidly and then reached a constant level, whereas the amount of the product with 3'-phosphate terminus ($P3'\text{P}$) increased linearly with time after a short lag (Fig. 4a). The kinetic behaviour strongly suggests a two-step mechanism where the $P2',3'\text{CP}$ is the reaction intermediate of the final reaction product, $P3'\text{P}$. Additional support for this mechanism was obtained by purifying the $P2',3'\text{CP}$ from LZ2, recombining it with the 24-mer product of the cleavage reaction, adding Pb^{2+} and following the time course of the reaction. The majority of $P2',3'\text{CP}$ was rapidly converted to $P3'\text{P}$ and a small amount was joined to the 24-mer RNA to reform the original 30-mer (Fig. 4b).

Although the data in Fig. 4b support the two-step mechanism, the small amount of 30-mer formed in the reaction left open the possibility that two reaction pathways could occur, one forming $P2',3'\text{CP}$ and a second forming $P3'\text{P}$ by direct nucleophilic attack of water on the 5'-3' phosphodiester linkage. In this view, the $P3'\text{P}$ formed in Fig. 4b would be the result of efficient ligation of $P2',3'\text{CP}$ to the 24-mer followed by direct attack of water on the 30-mer. This possibility was eliminated by phosphorylation of the 24-mer with polynucleotide kinase, making it incapable of undergoing ligation. When the 5' phosphorylated 24-mer RNA was combined with $P2',3'\text{CP}$, conversion of $P2',3'\text{CP}$ into $P3'\text{P}$ occurred at slower but significant rates (data not shown). This suggests that $P3'\text{P}$ can only be

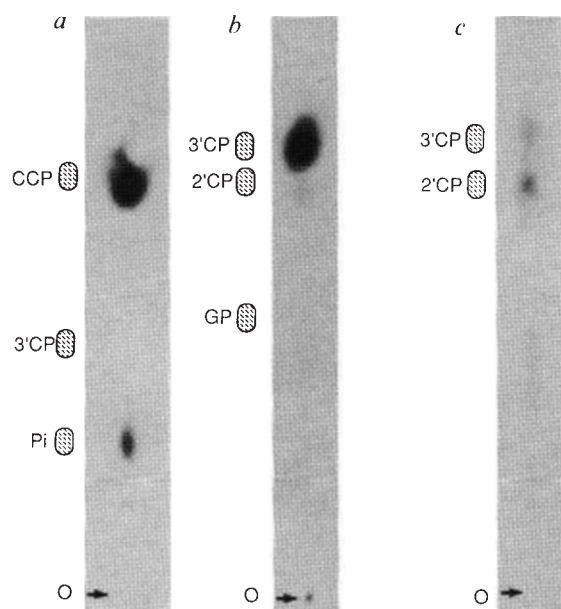


FIG. 3 Product identification by TLC. *a*, The upper band after treatment with nuclease P1 and calf-intestine alkaline phosphatase. *b*, The lower band after treatment with spleen phosphodiesterase II. *c*, The upper band was first incubated with Pb^{2+} and then treated with phosphodiesterase II.

METHODS. LZ3 was prepared by *in vitro* transcription with [α - ^{32}P]GTP. The gel-purified RNA was dephosphorylated with alkaline phosphate, phenol-extracted and ethanol-precipitated. Pb^{2+} cleavage was done in 15 mM MOPS, pH 7.0, 10 mM $MgCl_2$ and 0.2 mM lead acetate for 4 min at 25 °C and the two 5' cleavage products were purified from high-resolution denaturing gels. About 1 pmol of the slow-migrating oligomer in a 5 μ l reaction mixture containing 0.2 U nuclease P1, 0.3 U calf-intestine alkaline phosphatase and 50 mM ammonium acetate, pH 7.0 (*a*), and about 1 pmol of the fast-migrating oligomer in a 5 μ l reaction mixture containing 0.03 U phosphodiesterase II and 10 mM sodium phosphate, 0.5 mM EDTA, pH 6.5 (*b*), was incubated for 30 min at 37 °C. The reaction mixtures were then spotted on a cellulose polyethyleneimine TLC plate and developed with 1 M ammonium acetate. For *b* and *c*, a piece of Whatman 3 mm filter paper was attached to the top of the TLC plate to permit longer development time. Positions of non-radioactive markers are indicated on each side of the autoradiogram.

formed from $P2',3'CP$ and the direct hydrolysis pathway does not occur.

Specificity and rate enhancement

Pb^{2+} promotes rapid cleavage of unstructured RNA^{12,13}. This effect is believed to be due to second-order random collision of RNA with $Pb(OH)^+$. The relatively high rates of these non-specific reactions can be measured accurately to allow quantitative assessment of the catalytic efficiency of this Pb^{2+} ribozyme.

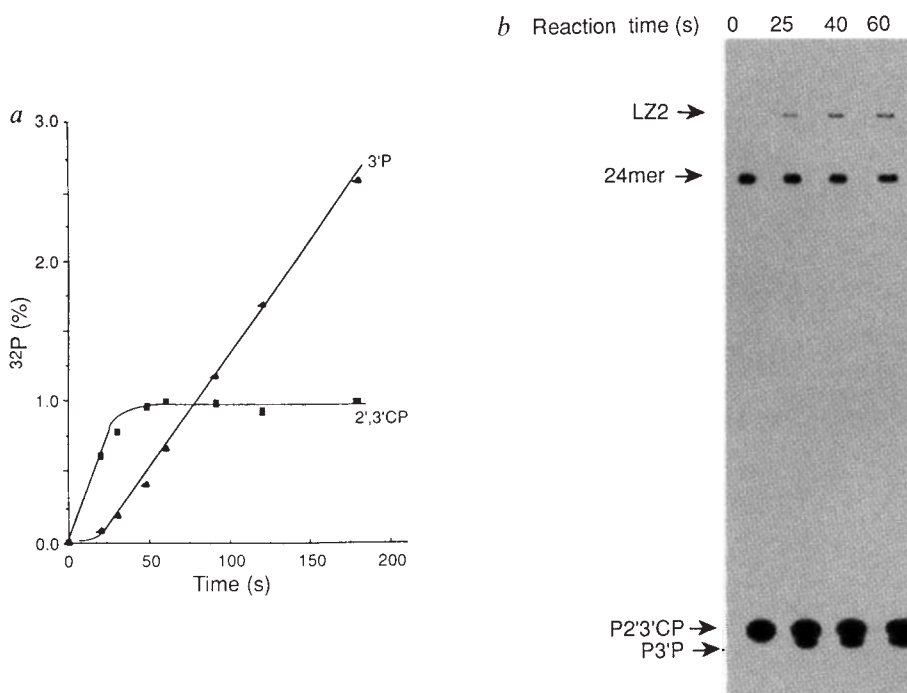
The pseudo-first-order rate of the autolytic cleavage of LZ2 by Pb^{2+} was compared to nonspecific Pb^{2+} cleavage of an RNA consisting of the 5' 14 nucleotides of LZ2. The observed rate constants, as measured by the disappearance of the full-length RNA molecules, showed a rate enhancement of 84-fold (0.24 min^{-1} against 0.00277 min^{-1} at 10 mM Mg^{2+} , 0.05 mM Pb^{2+} , pH 7.0 at 25 °C). But the specificity of the cleavage sites differed significantly between both RNA molecules. In the reaction of LZ2, cleavage at a single site was observed, whereas cleavage of the 14-nucleotide RNA was nonspecific. If we assume that all 13 5',3'-phosphodiester bonds within the 14-mer RNA were roughly equally susceptible to cleavage, the rate enhancement of the cleavage at the specific site can be calculated as $84 \times 13 = 1,100$ -fold.

For the specific hydrolysis of the 2',3'-cyclic phosphate to 3'-monophosphate, pseudo-first-order kinetics were measured using a large excess of the 24-mer RNA product of LZ2 cleavage (Fig. 4*b*) over the labelled substrate, $P2',3'CP$. For random hydrolysis, the first order kinetics of $P2',3'CP$ was measured in the absence of the 24-mer RNA. The rate enhancement of the 2',3'-cyclic phosphate hydrolysis was 24-fold (0.43 min^{-1} against 0.018 min^{-1} at 2.5 mM Mg^{2+} , 0.05 mM Pb^{2+} , pH 7.0 at 25 °C). More importantly, the reaction catalysed by the 24-mer produced almost exclusively 3'-monophosphate (Fig. 3*b*), whereas product identification by thin-layer chromatography of the non-specific reaction showed a mixture of 2'- and 3'-monophosphates of roughly equal amounts (Fig. 3*c*). This represents a rate enhancement of $24 \times 2 = 48$ -fold for the hydrolysis of the 2',3'-cyclic phosphate.

Discussion

We have described a Pb^{2+} -dependent catalytic RNA of unusually small size and simple secondary structure. The number

FIG. 4 *a*, Kinetics of 5'- ^{32}P -labelled LZ1 cleavage. Pb^{2+} cleavage was done at 25 °C in 15 mM MOPS, pH 7.0, 10 mM $MgCl_2$ and 0.2 mM lead acetate. At specified time points, aliquots of the reaction mixture were taken, the cleavage reaction stopped by addition of an equal volume of 9 M urea, 50 mM EDTA, and analysed on high-resolution denaturing gels containing 7 M urea. *b*, Direct conversion of $P2',3'CP$ to $P3'P$. 5'- ^{32}P -labelled $P2',3'CP$ and the α - ^{32}P -G-labelled 24-mer cleavage product of LZ2 were gel-purified and combined at 0.2 μ M $P2',3'CP$ and 1.2 μ M 24-mer. The conversion was done in 15 mM MOPS, pH 7.0, 2.5 mM $MgCl_2$ and 0.05 mM lead acetate, 25 °C. The reaction time in seconds is shown on top of the gel and the ligation product is indicated as LZ2.



of essential nucleotides is ten or less. Because both transesterification and subsequent hydrolysis reactions occur highly specifically, tertiary folding of the internal loop is likely to be required. Hydrolysis of the 2',3'-cyclic phosphate to 3'-monophosphate is not a property of all Pb^{2+} cleavage reactions. When the 5' Pb^{2+} cleavage product of yeast tRNA^{Phe} was analysed by thin-layer chromatography, no evidence of catalytic hydrolysis of the 2',3'-cyclic phosphate was found (data not shown). In a different reaction mechanism, specific hydrolysis of 5',3'-phosphodiester bond to 5'-monophosphate and 3'-hydroxyl termini by the *Tetrahymena* ribozyme^{14,15} and RNase P (refs 2, 16) has been observed.

The mechanism of the first reaction step for the Pb^{2+} ribozyme is likely to be similar to the Pb^{2+} cleavage of the yeast tRNA^{Phe}, where a Pb^{2+} specifically bound by nucleotide bases U59 and C60 extracts the proton from the 2'-OH of U17 to allow nucleophilic attack of the 2' oxyanion at the adjacent phosphodiester bond⁶. The first reaction step is similar to hammerhead, hairpin and hepatitis δ ribozymes in that cleavage is site-specific and produces 2',3'-cyclic phosphate⁹⁻¹¹. But ham-

merhead ribozyme can use several different metal ions³, whereas in our Pb^{2+} ribozyme, Pb^{2+} cannot be replaced by Zn^{2+} or Mn^{2+} (data not shown). Similarly, Pb^{2+} cannot substitute for Mg^{2+} in the hammerhead ribozyme³. By contrast, because of the much higher uncatalysed rate constant by Pb^{2+} , the rate enhancement of this Pb^{2+} ribozyme is only three orders of magnitude compared to the rate stimulation of nearly seven orders of magnitude above background by these Mg^{2+} ribozymes¹⁵.

Because of the lack of titratable functional groups near physiological pH, RNA molecules are intrinsically poor general acid-base catalysts¹⁴. In some cases, tertiary folding of RNA molecules can cause strong shifts in the proton association constants of functional groups of the nucleotide bases¹⁷. But the hydrophilic and negatively charged RNA molecules are well suited for binding metal ions which can then act as functional equivalents of the ionizable amino-acid side chains. Thus, the potential catalytic power of RNAs as general acid-base catalysts may be elucidated by isolation and design of specific RNA metal-ion binding domains. □

Received 27 April; accepted 29 June 1992.

- Cech, T. R. *Rev. Biochem.* **59**, 543-568 (1990).
- Pace, N. & Smith, D. *J. biol. Chem.* **265**, 3587-3590 (1990).
- Dahm, S. C. & Uhlenbeck, O. C. *Biochemistry* **30**, 9464-9469 (1991).
- Rubin, J. R. & Sundaralingam, M. *J. biomolec. Struct. Dyn.* **1**, 639-646 (1983).
- Brown, R., Hingerty, B. E., Dewan, J. & Klug, A. *Nature* **303**, 543-546 (1983).
- Brown, R., Dewan, J. & Klug, A. *Biochemistry* **24**, 4785-4801 (1985).
- Behlen, L. S., Sampson, J. R., DiRenzo, A. B. & Uhlenbeck, O. C. *Biochemistry* **29**, 2515-2523 (1990).
- Pan, T. & Uhlenbeck, O. C. *Biochemistry* **31**, 3887-3895 (1992).
- Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. *Science* **231**, 1577-1580 (1986).
- Uhlenbeck, O. C. *Nature* **328**, 596-600 (1987).
- Perrotta, A. T. & Been, M. D. *Biochemistry* **31**, 16-21 (1992).
- Farkas, W. R. *Biochim. biophys. Acta* **155**, 401-409 (1968).

- Sundaralingam, M., Rubin, J. R. & Cannon, J. F. *J. quant. Chem. quant. Biol. Symp.* **11**, 355-366 (1984).
- Cech, T. R. & Bass, B. L. *Rev. Biochem.* **55**, 599-629 (1986).
- Cech, T. R. *Science* **236**, 1532-1539 (1987).
- Altman, S. *Adv. Enzym. relat. Areas molec. Biol.* **62**, 1-36 (1989).
- Kao, T. H. & Crothers, D. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3360-3364 (1980).
- Forster, A. & Symons, R. H. *Cell* **50**, 9-16 (1987).
- Milligan, J. F. & Uhlenbeck, O. C. *Meth. Enzym.* **180**, 51-64 (1989).

ACKNOWLEDGEMENTS. We thank T. Cech, M. Fedor, J. Doudna and P. Legault for their comments on the manuscript. This work was supported by grants from the NIH. T.P. is supported by a postdoctoral fellowship from the Damon Runyon-Walter Winchell Cancer Research Fund. We also thank the W. M. Keck Foundation for their generous support of RNA science on the Boulder campus.

Identification of V1017 Sgr as a cataclysmic variable binary system with unusually long period

Kazuhiro Sekiguchi

South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa

CLASSICAL and dwarf novae are both binary star systems in which mass is transferred onto a white-dwarf primary from a red-dwarf secondary that overflows its Roche lobe. In classical novae, mass accumulates on the surface of the white dwarf and erupts in a thermonuclear outburst every ten thousand years or so, whereas dwarf novae show less intense and more frequent outbursts, on a timescale of months, due to brightening of an accretion disk. V1017 Sgr, the remnant of nova Sagittarii 1919, has shown outbursts of both classical and dwarf nova type during this century. It has been described as a recurrent nova^{1,2} or as a symbiotic star^{2,3}. Unlike most novae it has a more massive evolved secondary, of type G5 III (ref. 4), and its relation to other nova types has been obscure. Here I report a preliminary spectroscopic orbit determination which yields a period of 5.7 days. This is much longer than the typical periods of cataclysmic binaries, which are generally less than 0.5 day. I suggest that V1017 Sgr, along with a few other systems of unusually long period such as BV Cen and GK Per, may form a distinct class of cataclysmic variables, characterized by their larger, more evolved secondaries.

V1017 Sgr is an unusual member of the class of cataclysmic variable (CV) stars which has produced recorded outbursts in 1901, 1919, 1973 and 1991. Three of these outbursts (1901, 1973

and 1991) were very similar, with a slow symmetric rise and decay by ~ 3 mag in ~ 200 day. Although almost all dwarf novae (DNe) show much faster rise times (~ 1 -2 day), these three outbursts are most likely to be accretion-powered DN-type events. In contrast, the 1919 outburst was more energetic, reaching a peak magnitude of $m_{\text{pg}} \leq 7.2$ (ref. 5) from its mean magnitude at minimum $V = 13.59 \pm 0.07$ (ref. 6) and declining more slowly. Unfortunately, no spectrum exists for the 1919 outburst, but its photometric pattern suggests a classical nova (CN) outburst. Following this, no DN-type outburst was observed until 1973. This 54-yr interval (3×18 yr) and the 18-yr interval between the 1973 and the 1991 DN-type outbursts suggest that the true timescale for recurrence of DN-type outbursts in the system may be fairly regular ($\tau_{\text{rec}} \approx 18$ yr). The CN outburst of 1919 (18 years since the 1901 DN-type outburst) may have been preceded by a DN-type outburst that dumped enough material to ignite the thermonuclear outburst. After the 1919 CN outburst, something must have suppressed the two DN-type outbursts expected in 1937 and 1955. A possible explanation is that the disk material was heated by the hot white dwarf so that the disk was kept above the critical temperature. By 1973 the white dwarf had cooled to the point where the disk was again below the critical temperature and DN-type outbursts could start again. This makes the system more likely to be a CN with DN-type eruptions before and after the CN outburst. Only a few systems are known to show such behaviour⁶⁻⁹ and they may provide clues to the question of whether classical and dwarf novae are in fact the same systems undergoing cyclic evolution. Despite its potential importance, the structure and evolutionary status of V1017 Sgr are not currently understood. Here I report the preliminary results of a spectroscopic orbit determination of V1017 Sgr and discuss the implications for the nature of this unusual object.

TABLE 1 Radial velocities from absorption lines

HJD 2,440,000+	Radial velocity (km s ⁻¹)	Mean error (km s ⁻¹)
8477.4023	89.1	3.0
8478.3916	0.2	25.0
8480.4072	-74.0	0.8
8481.3408	-11.5	11.8
8510.2910	4.2	0.8
8511.3486	88.2	1.5
8512.3652	75.4	11.9
8513.2754	14.4	7.0
8514.3760	-58.8	1.2

The spectra were all taken with the grating spectrograph and intensified photon-counting Reticon detector attached to the 1.9-m reflector of the South African Astronomical Observatory at Sutherland. All the spectra have a dispersion of $30 \text{ \AA mm}^{-1}$ (corresponding to $\sim 1.2 \text{ \AA}$ full width at half maximum) and cover the wavelength region from 4,050 to 4,450 $\text{\AA}$. Comparison of the average spectrum of V1017 Sgr with that of a G5 III star (Fig. 1) confirms that the secondary should be classified as a G giant⁴. Three radial-velocity standard stars, HD157457 (G8 III), HD168454 (K2.5 III) and HD171391 (G8 III) were observed before and after each nova observation. I used a cross-correlation technique¹⁰ to determine the radial velocities of the underlying G star (Table 1).

I have used Fourier power spectra to determine the period of the radial velocity shifts in the range 0.5–20.0 day. The highest peaks occur at periods of 0.851, 1.212 and 5.714 day, all giving acceptable fits to the data. But observations identifying the Roche-lobe-filling secondary as a G5 giant star point to a longer period ($\sim 2\text{--}20$ day)⁶. Furthermore, the radial velocity curve (Fig. 2, upper curve) suggests an orbital period of a few days. I adopted a 5.714-day orbital period with the epoch of superior conjunction of the G star at HJD (heliocentric Julian day) 2,448,478.983, but the other two periods must also be considered possible. Obviously, more data would clarify this and allow

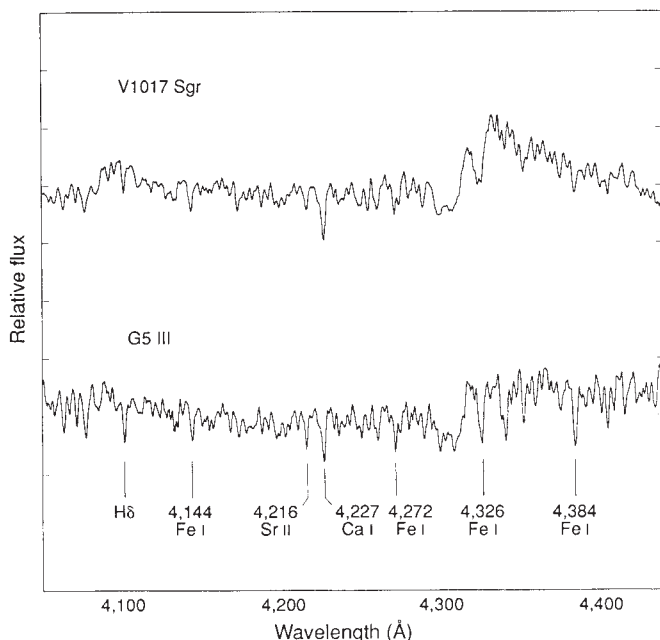


FIG. 1 Mean spectrum of V1017 Sgr (upper curve) in the wavelength region 4,050–4,450 $\text{\AA}$ compared to that of a G5 III star BS-332 (lower curve).

determination of the orbital elements. If a long orbital period of ~ 5.7 days is confirmed, it would be the longest orbital period known for a CN remnant.

Apart from recurrent novae, only the following five CVs are known to have an orbital period longer than 0.5 days. GK Per (CN remnant; Nova Persei 1901) has an orbital period of $P = 1.996$ days and a secondary type K0 IV and mass ~ 0.25 solar masses ($M_{\odot}$)¹¹; BV Cen (DN) has $P = 0.611$ days and a G5–8 IV–V ($m \approx 0.9 M_{\odot}$) secondary¹²; V841 Oph (CN remnant; Nova Ophiuchus 1848) has $P = 0.6$ days¹³; DI Lac (CN remnant; Nova Lacerta 1910) has $P = 0.54$ days¹⁴; and V Sge (nova-like variable) has $P = 0.51$ days¹⁴. Two of these, BV Cen and GK Per, also show DN-type outbursts with a slow symmetric rise to and decay from the maximum. Furthermore, both systems are transient hard X-ray sources¹⁵. This fact led Menzies *et al.*¹⁶ to suggest that BV Cen was an undetected CN remnant of GK Per type. Spectra observed during the 1991 outburst of V1017 Sgr¹⁷ are similar to those of GK Per¹⁸ and BV Cen (K.S. *et al.*, manuscript in preparation), having a blue continuum and strong He II, 4,640 $\text{\AA}$, He I and Balmer emission lines. The characteristics of these three DN-type outbursts are generally similar. Although the development of the outburst light curve is slower for V1017 Sgr, with rise times of ~ 100 days, than for GK Per and BV Cen (~ 25 and ~ 15 days, respectively), and its recurrence timescale ($\tau_{\text{rec}} \approx 18$ yr) is much longer than those of GK Per and BV Cen ($\sim 1,000$ and ~ 150 days), these timescales correlate with the system's orbital period (the size of the accretion disk). Although V841 Oph is an X-ray source¹⁹, it shows a sinusoidal optical variation, which is not characteristic

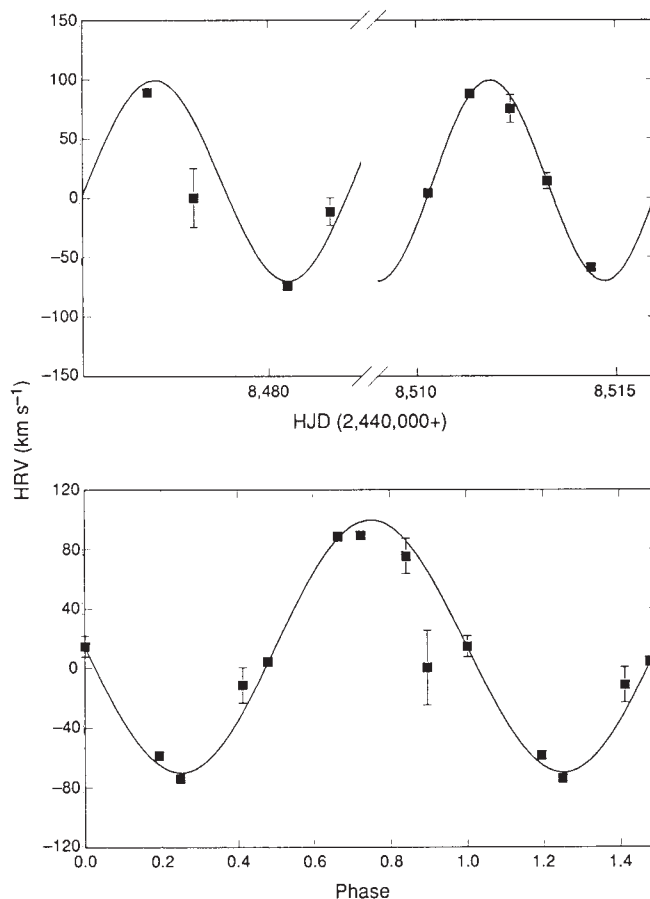


FIG. 2 Observed radial velocities are plotted with Fourier curve for the orbital period 5.714 day (upper panel). The radial velocity variation is shown as a function of phase of the absorption lines (lower panel). The smooth curve represents Fourier curve fits to the data. HRV, heliocentric radial velocity (corrected for motion of Earth in orbit).

of DN-type outburst, with amplitude of 0.4 mag and a period of ~ 40 days (ref. 20). V841 Oph therefore does not belong to the class of long-period hybrid classical/dwarf novae. Further observations are needed for DI Lac and V Sge, but the minimum spectrum of DI Lac closely resembles that of V841 Oph²¹.

In conclusion, I suggest that V1017 Sgr has the longest known orbital period of a population of CVs characterized by thermonuclear (CN) outbursts, accretion (DN-type) outbursts with a symmetric light curve and slow rise to the maximum, a long orbital period $P > 0.5$ days and a hard X-ray source. The detailed structure of these systems and their evolutionary status are not yet understood. Their long orbital periods imply that they have a large accretion disk and an evolved secondary. The accretion in these systems may be driven by nuclear evolution. It is possible

that these systems represent a sequence of evolution of the same type of object. The observed differences between these systems and the recurrent novae of the U Sco type, whose orbital periods are ~ 0.5 –2.0 days, may simply be due to a different rate of mass transfer. A possible connection between these systems and the shorter orbital period CN remnants with DN-type outburst activity (for example WY Sge) must be investigated. A recent model for the DN-type outbursts of GK Per²² gives a symmetric outburst by assuming a high rate of mass transfer with the heating wave starting in the inner disk and propagating outward (this suggests that the outburst cannot result from a mass-transfer instability). Attempts to model BV Cen and V1017 Sgr in the same way and compare these three systems would be rewarding. $\square$

Received 2 March; accepted 22 June 1992.

- Payne-Gaposchkin, C. *The Galactic Novae* (North-Holland, Amsterdam, 1957).
- Kenyon, S. J. *The Symbiotic Stars* (Cambridge Univ. Press, 1986).
- Kukarkin, B. V. *2nd Supplement to 3rd Edn of the General Catalogue of Variable Stars* (Nauka, Moscow, 1974).
- Kraft, R. P. *Astrophys. J.* **139**, 457–475 (1964).
- McLaughlin, D. B. *Publ. astr. Soc. Pacif.* **58**, 46–52 (1946).
- Webbink, R. F. et al. *Astrophys. J.* **314**, 653–672 (1987).
- Vogt, N. in *Classical Novae* (eds Bode, M. F. & Evans, A.) 225–247 (Wiley, Chichester, 1989).
- Livio, M. *Comments Astrophys.* **12**, 87–97 (1987).
- Livio, M. in *The Symbiotic Phenomenon, Proc. Int. Astr. Un. Colloq. 103* (eds Mikolajewska, J. et al.) 323–334 (Kluwer, Dordrecht, 1988).
- Tony, J. & Davis, M. *Astr. J.* **84**, 1511–1525 (1979).
- Crampton, D., Cowley, A. P. & Fisher, W. A. *Astrophys. J.* **300**, 788–793 (1986).
- Gilliland, R. *Astrophys. J.* **263**, 302–311 (1982).

- Bianchini, A., Friedjung, M. & Sabbadin, F. in *Physics of Classical Novae, int. astr. Un. Colloq. 122* (eds Cassatella, A. & Viotta, R.) 61–62 (Springer, Berlin, 1990).
- Ritter, H. *Astr. Astrophys. Suppl.* **57**, 385–418 (1984).
- Cordova, F. A. & Mason, K. O. *Mont. Not. R. astr. Soc.* **206**, 879–897 (1984).
- Menzies, J. W., O'Donoghue, D. & Warner, B. *Astrophys. Space Sci.* **122**, 73–80 (1986).
- Sekiguchi, K. & Latham, D. W. *Int. astr. Un. Circ.* **5360** (1991).
- Bianchini, A., Sabbadin, F. & Hamzaoglu, E. *Astr. Astrophys.* **106**, 176–178 (1982).
- Becker, R. H. & Marshall, F. E. *Astrophys. J.* **244**, L93–L95 (1981).
- Shara, M. M., Potter, M. & Shara, D. J. *Publ. astr. Soc. Pacif.* **101**, 985–990 (1989).
- Williams, G. *Astrophys. J. Suppl.* **53**, 523–551 (1983).
- Kim, S.-W., Wheeler, J. C. & Mineshige, S. *Astrophys. J.* **384**, 269–283 (1992).

ACKNOWLEDGEMENTS. I thank B. Warner for suggesting why DN-type outbursts did not occur in 1937 and 1955, and for discussions. I also thank M. W. Feast, J. W. Menzies and R. F. Webbink for comments and suggestions.

Observation of an electronic bound state above a potential well

Federico Capasso, Carlo Sirtori, Jerome Faist, Deborah L. Sivco, Sung-Nee G. Chu & Alfred Y. Cho

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

SHORTLY after the birth of quantum mechanics, von Neumann and Wigner made the remarkable proposal¹ that certain spatially oscillating attractive potentials could support bound states at energies above the potential barriers (that is, spatially confined states within the continuum) by means of diffractive interference. Because of their unusual geometry, such potentials were regarded as mathematical curiosities^{2,3}, although more recently it has been suggested that they might be found in certain atomic and molecular systems^{4,5}. Following the observation of discrete electronic states in ultra-thin semiconductor layered structures^{6,7} (for example, in quantum wells), Stillinger⁸ and Herrick⁹ proposed that superlattices might be used to construct potentials supporting these 'positive energy' bound states. Here we report direct evidence of such states in semiconductor heterostructures grown by molecular-beam epitaxy¹⁰. Infrared absorption measurements reveal a narrow, isolated transition from a bound state within a quantum well to a bound state at an energy greater than the barrier height; this state is spatially localized by Bragg reflections.

Figure 1 shows the conduction-band energy diagrams of our layered semiconductor structures. Such diagrams¹¹ are one-dimensional representations of the energy levels along the direction of growth (normal to the layers). Because of the atomic flatness and uniform thickness of layers grown by molecular-beam epitaxy (MBE)¹⁰, we can assume translational invariance parallel to the layers. As a result the states have plane-wave-like dispersion parallel to the layers. The discrete energy levels in Fig. 1 are therefore the bottom of energy sub-bands, and the corresponding states can be considered as wavefunctions in the direction normal to the layers (or equivalently at zero lateral wave vector), because of the decoupling between normal and

lateral degrees of freedom. The states are calculated by solving Schrödinger's equation using Bastard's envelope function approximation¹¹. For the parameters of our $\text{Al}_{0.48}\text{In}_{0.52}\text{As}/\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ heterostructures we used $\Delta E_c = 0.50$ eV for the barrier height, $m_e^*(\text{GaInAs}) = 0.043 m_0$, $m_e^*(\text{AlInAs}) = 0.07 m_0$ for the effective masses and $\gamma = 1.03 \times 10^{-18} \text{ m}^2$ for the nonparabolicity coefficient. Deviations from parabolicity were taken into account using the method of ref. 12.

Consider a thin quantum well with thick barriers, designed to have a single bound state of energy E_1 . At energies greater than the barrier height, there are only scattering states. At certain energies, corresponding to a semi-integer number of de Broglie wavelengths across the well, one finds enhanced transmission resonances¹¹. Although at these energies the electron amplitude in the well layer is enhanced, the corresponding states remain extended. Figure 1a represents the energy diagram of a series of such wells, consisting of 32 Å layers of GaInAs separated by 150-Å-thick AlInAs barriers. Our reference sample contains 20 of these wells doped n-type with silicon ($3 \times 10^{11} \text{ cm}^{-2}$). Note that at energies $> \Delta E_c$, energy bands called minibands are formed from the coupling between wells¹¹. On the other hand the confined states of the wells (E_1) do not form a miniband because of localization effects associated with unavoidable scattering and imperfections, given their negligible tunnelling amplitude.

To construct bound states (that is, states with square-integrable wavefunctions) at energies greater than the barrier height, we replace the barriers with AlInAs/GaInAs superlattice barriers of equal height. The superlattice is designed so that Bragg reflections spatially localize the state corresponding to the first continuum resonance of the well. The energy of this resonance, E_2 , is found from¹¹

$$k_W L_W = \pi \quad (1)$$

L_W is the well thickness and $k_W = \sqrt{2m_w^*(E)E/\hbar^2}$ is the electron wavenumber in the well material (GaInAs), where E is the electron energy and $m_w^*(E)$ the energy-dependent electron effective mass defined in ref. 12. From equation (1) we obtain $E_2 = 560$ meV for $L_W = 32$ Å. All the waves partially reflected by the superlattice interfaces must constructively interfere to form a bound state at the energy E_2 , within the superlattice minigap (Fig. 1b). To achieve this, the thickness of the

superlattice wells (d_W) and barriers (d_B) are chosen to be a quarter of the corresponding de Broglie wavelength at the energy of the first transmission resonance E_2 determined by equation (1)

$$k_W d_W = \frac{\pi}{2} \quad (2)$$

$$k_B d_B = \frac{\pi}{2} \quad (3)$$

where $k_B = \sqrt{2m_B^*(E_2)(E_2 - \Delta E_c)/\hbar^2}$ is the wavenumber in the barrier material at the energy E_2 ; $m_B^*(E_2)$ is the effective mass at the energy E_2 in the barrier material. As k_W has the same value in equations (1) and (2), it follows immediately that $d_W = L_W/2 = 16 \text{ \AA}$. From the value of E_2 , we then calculate k_B ; using equation (3) one then finds $d_B = 39 \text{ \AA}$. The effect of cladding the wells with Bragg mirrors can be viewed as the equivalent of using high-reflectivity (ideally unity) quarter-wave dielectric

stacks to produce a Fabry-Perot optical cavity¹³. Because of the finite thickness of the Bragg reflectors in a real structure, the state at energy E_2 is, strictly speaking, a quasi-bound state with an extremely narrow width. The maximum useful thickness of the Bragg reflectors is set by the coherence length l_c of an electron in the excited state E_2 . Transport experiments at 10 K in AlInAs/GaInAs superlattices of similar layer thicknesses¹⁴ gave $l_c \approx 300 \text{ \AA}$. We chose, therefore, Bragg reflectors with six periods designed as above. The energy level E_2 can be also understood as a deep level in the superlattice bandgap, arising from the introduction of an 'artificial defect' (the central well in Fig. 1b) in a periodic superlattice structure.

Our structure with Bragg reflectors consists of 22 six-period undoped AlInAs/GaInAs superlattices, designed as above, separated by twenty 32-Å GaInAs wells doped n-type with silicon to give an electron sheet density of $3 \times 10^{11} \text{ cm}^{-2}$ necessary for infrared absorption between sub-bands. The Fermi energy is $\sim 15 \text{ meV}$. Calculations show that band bending in the doped wells and in the barrier regions is negligible. In this sample and in the reference structure, the growth initiates and terminates with an n-type doped GaInAs buffer layer. Figure 1c is a transmission electron micrograph of a cleaved cross-section, showing a GaInAs well sandwiched between the two superlattice barriers acting as Bragg reflectors.

For the experiments, we processed our samples in multipass (six) 45° wedge waveguides. This geometry allows us to couple in linearly polarized radiation with a large component of the polarization normal to the layers, as required by the inter-sub-band absorption selection rules¹⁵. Absorption measurements were made with a Nicolet system 800 Fourier transform infrared spectrometer. The samples were cooled in a Helitran flow dewar. Figure 2 shows the absorption spectra at 10 K of the multiple quantum well structures with and without Bragg reflectors. In the spectrum of the reference structure, the low-energy cutoff ($\hbar\omega \approx \Delta E_c - E_1 = 0.3 \text{ eV}$, where ω is angular frequency) corresponds to the onset of the classical continuum (Fig. 1a). The

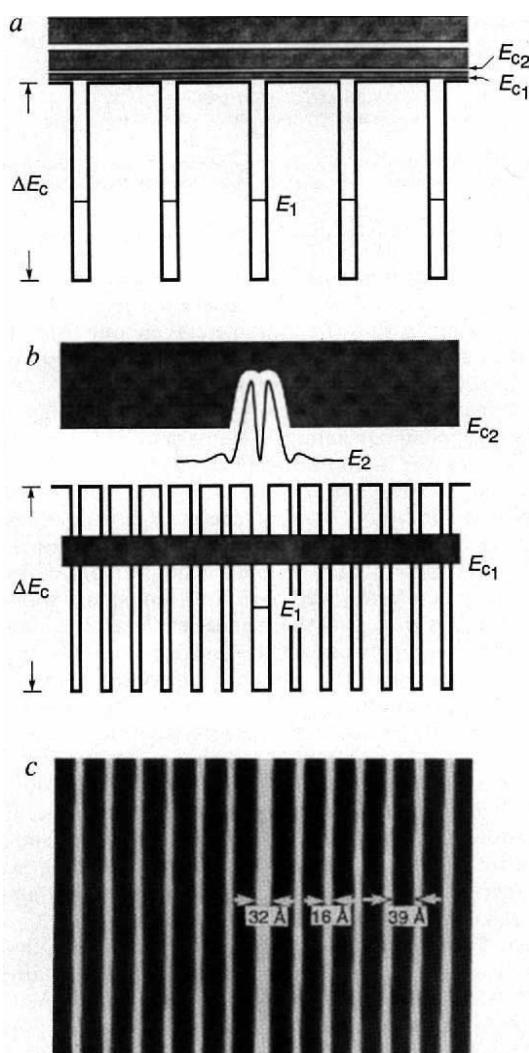


FIG. 1 Conduction band diagrams of AlInAs/GaInAs heterostructures. All energies are measured from the bottom of the wells; E_{c1} and E_{c2} indicate conduction miniband edges. a, GaInAs quantum wells (32 Å) with AlInAs barriers (reference sample). Shown are the ground state of the well ($E_1 = 204 \text{ meV}$) and the minibands in the continuum. b, GaInAs quantum well (32 Å) with Bragg reflector barriers produced by the AlInAs/GaInAs superlattice. Shown is probability density $|\psi|^2$ of the bound state ($E_2 = 560 \text{ meV}$) localized above the well in the superlattice minigap (266 meV). c, Transmission electron micrograph of a cleaved cross-section of the structure with superlattice Bragg reflectors, showing the atomic abruptness of the interfaces.

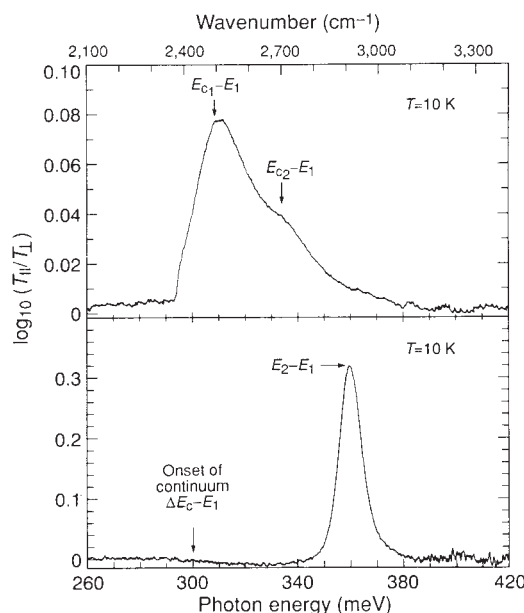


FIG. 2 Absorption spectra at room temperature for the reference structure of Fig. 1a (top) and the structure with superlattice Bragg reflectors of Fig. 1b (bottom). The absorbance for polarization parallel to the plane of the layers ($-\log_{10} T_{\parallel}$) is subtracted from the absorbance normal to the plane of the layers ($-\log_{10} T_{\perp}$), to remove the contribution of free carrier absorption in the buffer layers. The resulting quantity is proportional to the inter-sub-band absorption coefficient. The transition to the confined state above the well (E_2 in Fig. 1b) corresponds to the peak at 360 meV, in excellent agreement with the calculated value ($E_2 - E_1$).

energy of the absorption peak is close to the energy difference between the edge of the lower miniband and the bound state of the well, $E_{c1} - E_1$, very near the onset of the continuum, as expected from the calculated spectrum and band diagram. The shoulder at $E_{c2} - E_1$ represents the onset of absorption by the second miniband, in good agreement with the theoretical value. Absorption to higher bands is not observed because the matrix elements are so small. In the multiwell structure with Bragg reflectors we instead observe an isolated narrow peak corresponding to a transition to a state at an energy above the barrier height ΔE_c . It is centred at an energy of 360 meV and has a full width at half maximum of 9.7 meV; its width at 300 K is 17.5 meV. The position of this peak agrees well with the calculations. The absorption peak is much narrower than the spectrum corresponding to the bound-state/extended-state transition in the reference sample. Its width reflects contributions from various mechanisms estimated as follows: broadening due to nonparabolicities (~ 2.5 meV), inter-sub-band scattering (~ 2 meV) and atomic-scale in-plane variations in thickness (~ 5 meV). From the area under the peak and the electron density in the wells, we estimate a transition matrix element $\langle z_{12} \rangle \approx 10$ Å consistent with the theoretical value of 10.5 Å. This corresponds to an oscillator strength $f = 0.5$. These high values of $\langle z_{12} \rangle$ and f are comparable to those calculated and measured for bound-state/bound-state transitions in conventional AlInAs/GaInAs quantum wells¹⁶. The width of the absorption peak at $\hbar\omega = 360$ meV is virtually equal to that of bound-state/bound-state transitions measured by us in conventional 55-Å GaInAs quantum wells with 300-Å AlInAs barriers doped to a comparable level. These wells were designed to have two bound states below the barrier height. Finally, absorption measurements in samples with one- and two-period quarter-wave ($\lambda/4$) stacks of identical layer thicknesses show that the position of the peak, as expected, does not shift and that the linewidth is reduced with increasing number of periods, in accordance with the increased localization of the final state.

In the structure of Fig. 1b, we observe at low energy an absorption band in the range 100–180 meV associated with photoexcitation of electrons from the ground state (E_1) to the low-energy miniband extending from 307 to 379 meV. This miniband arises from the tunnel coupling between the states of the wells in the $\lambda/4$ stacks.

The use of $\lambda/4$ stacks to enhance barrier reflectivity has been proposed for semiconductor lasers¹⁷ and demonstrated in tunnelling experiments¹⁸. Structures with states confined in a barrier by $\lambda/4$ stacks, similar to the one originally proposed⁸, have recently been considered¹⁹ and experimentally investigated²⁰. The latter states should be distinguished from resonant states with a large amplitude in the barrier layers recently investigated by some of us²¹ and other authors²².

This work illustrates how bandgap engineering²³ can be used to design almost arbitrary potentials so that wavefunctions and electronic properties can be tailored. We plan to investigate in detail the effect of electric and magnetic fields. Besides being of fundamental quantum mechanical interest the new states studied here should encourage further investigations of the analogies between electronics and optics^{24,25}. □

14. Beltram, F. *et al.* *Phys. Rev. Lett.* **64**, 3167–3170 (1990).
15. West, L. C. & Eglash, S. J. *Appl. Phys. Lett.* **46**, 1156–1158 (1985).
16. Levine, B. F. *et al.* *Appl. Phys. Lett.* **52**, 1481–1483 (1988).
17. Iga, K., Uenohara, H. & Koyama, F. *Electron Lett.* **22**, 1008–1010 (1986).
18. Takagi, T., Koyama, F. & Iga, K. *Appl. Phys. Lett.* **59**, 2877–2879 (1991).
19. Lenz, G. & Salzman, J. *Appl. Phys. Lett.* **56**, 871–873 (1990).
20. Brener, I. *et al.* *Quantum Electronics and Laser Science Conf. 1992 OSA Technical Digest Series*, Vol. 13, 216–217 (Optical Society of America, Washington DC, 1992).
21. Sirtori, C., Capasso, F., Sivco, D. L. & Cho, A. Y. *Appl. Phys. Lett.* **60**, 2678–2680 (1992).
22. Shen, H., Pollack, F. H. & Tsu, R. *Appl. Phys. Lett.* **57**, 13–15 (1990).
23. Capasso, F. *Science* **235**, 172–176 (1987).
24. *Analogies in Optics and Microelectronics* (eds van Hoeninger, W. & Lenstra, D.) (Kluwer, Dordrecht, 1990).
25. Gaylord, T. K. & Brennan, K. F. *J. appl. Phys.* **65**, 814–820 (1989).

ACKNOWLEDGEMENTS. We thank J. S. Weiner for bringing to our attention refs 1, 8 and 9. C.S. acknowledges financial support from Consorzio Industriale Technobiochip, Marciana, Italy.

Novel field-induced asymmetry in the remanent magnetization of the superconductor $\text{YBa}_2\text{Cu}_3\text{O}_7$

D. A. Brawner, N. P. Ong & Z. Z. Wang*

Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544, USA

* Lab 2M/CNRS, 196 avenue Henri Ravéra, 92220 Bagneux, France

THE dynamics of trapped vortices of magnetic flux in the high- T_c oxide superconductors have been much studied^{1–4}, but remain poorly understood; in particular, we have much to learn about vortex 'creep' at low temperatures. Using a scanning Hall microprobe to measure directly the spatial and temporal variation of the flux density in $\text{YBa}_2\text{Cu}_3\text{O}_7$ single crystals, we have uncovered a remarkable asymmetry in the magnetization profile, which is induced by the direction of the field (applied normal to the a - b crystal face). The external field imparts a rotation to the vortex motion, which strongly distorts the remanent profile away from the form predicted by conventional critical-state models⁵. This rotation was not anticipated and, to our knowledge, is not accounted for in existing theories.

A Hall microprobe, consisting of a 1,500-Å bismuth film with an active area $2 \times 4 \mu\text{m}^2$ and sensitive to changes in field of ± 0.3 Oe (D.A.B. and N.P.O., manuscript in preparation), is scanned at a constant height $15 \mu\text{m}$ above the a - b face of a $\text{YBa}_2\text{Cu}_3\text{O}_7$ (YBCO) crystal (Fig. 1, inset). Typically, one-dimensional scans along the x -axis (parallel to the shorter side of the a - b face) are done at a fixed value of y . The Hall probe measures the z -component of the local field induction $\mathbf{B}(x, y)$, associated with the spot (x, y) .

Remanent magnetization is trapped in a type II superconductor when the external field $\mathbf{H}$ is increased to a maximum value (larger than that required for complete flux penetration) and then returned to zero. In the Bean model⁵, the saturated remanence profile is nominally a triangle with slopes equal to the critical-state current density J_c . For a sample shaped like a rectangular plate, a generalization of the 'critical-state' assumption predicts that the remanence is shaped like a roof with the gradient everywhere equal to J_c . By symmetry, the vortex density $B(x, y)$ attains its maximum value on a straight line (the 'ridge') that is parallel to the long side of the sample and passes through the centre of the crystal, for either field direction.

Our experiment shows that this simple picture is invalid in $\text{YBa}_2\text{Cu}_3\text{O}_7$ at low temperatures T . The shape of the remanence profile is skewed in a direction dictated by the field direction. Figure 1 compares the positive and negative remanence profiles for two scans A and B parallel to x at 4.2 K (inset). For the scan along the line A, the remanent profile created in positive field 'leans' to the right whereas the profile (negative in sign) created in negative field leans to the left. For scans at b , however, the two profiles lean in the opposite directions. To clarify how

Received 5 May; accepted 21 July 1992.

1. von Neumann, J. & Wigner, E. *Phys. Z.* **30**, 465–467 (1929).
2. Simon, B. *Commun. pure appl. Math.* **22**, 531–538 (1969).
3. Weidmann, J. *Math. Z.* **98**, 268–302 (1967).
4. Stillinger, F. H. & Weber, T. A. *Phys. Rev. A* **10**, 1122–1130 (1974).
5. Stillinger, F. H. & Herrick, D. R. *Phys. Rev. A* **11**, 446–454 (1975).
6. Esaki, L. & Chang, L. L. *Phys. Rev. Lett.* **33**, 495–498 (1974).
7. Dingle, R., Wiegmann, W. & Henry, C. H. *Phys. Rev. Lett.* **33**, 827–830 (1974).
8. Stillinger, F. H. *Physica B* **85**, 270–276 (1977).
9. Herrick, D. R. *Physica B* **85**, 44–50 (1977).
10. Cho, A. Y. *J. Cryst. Growth* **111**, 1–13 (1991).
11. Bastard, G. *Wave Mechanics Applied to Heterostructures* (Editions de Physique, Paris, 1990).
12. Nelson, D. F., Miller, R. C. & Kleinman, D. A. *Phys. Rev. B* **35**, 7770–7773 (1981).
13. Macleod, H. A. *Thin Film Optical Filters* 2nd Edn, Ch. 7, 244 (Macmillan, New York, 1986).

the asymmetry depends on y , we show six scans in Fig. 2 through different sections of the crystals (Fig. 2A–F). Clearly, rather than being parallel to y , the ridge roughly coincides with a diagonal of the a – b face. If $\mathbf{H}$ is reversed, the ridge aligns with the other diagonal. Thus, the external field causes the ridge to rotate by some angle Ω_{out} . Associating an axial vector $\hat{\Omega}_{\text{out}}$ with the sense of the rotation, we find that $\hat{\Omega}_{\text{out}}$ is invariably antiparallel to $\mathbf{H}$. A condition for observing the rotation is that J_c should be sufficiently large. When we warm YBCO above ~ 20 K, the asymmetry rapidly becomes harder to discern (Fig. 3) (J_c decreases exponentially from $3 \times 10^6 \text{ A cm}^{-2}$ at 4.2 K to $4 \times 10^5 \text{ A cm}^{-2}$ at 25 K). In $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$ crystals, which have smaller J_c ($2 \times 10^5 \text{ A cm}^{-2}$ at 4.2 K), we do not observe any asymmetry.

We have carefully excluded an extrinsic cause for the asymmetry. From a series of experiments in which the sample was cooled initially either in zero field or in an intense field of either direction, we find that the sign and magnitude of the asymmetry are independent of the sample's magnetization or temperature history. Altogether, we have investigated seven crystals with different aspect ratios and thicknesses. All displayed the same asymmetry, with $\hat{\Omega}_{\text{out}} \parallel -\mathbf{H}$. The crystals, grown in a BaCuO_{2+y} flux, are of high quality with sharp resistive transitions (width of 0.3 K) at 91–92 K, and in-plane resistivities (a good indicator of oxygen homogeneity) from 50 to 70 $\mu\Omega \text{ cm}$ near 100 K (refs 6, 7). In addition to the crystals, we also observe the asymmetry in melt-textured (c -axis aligned) polycrystalline samples of YBCO. Cutting these latter samples into smaller pieces did not affect the rotation. We also did tests to eliminate microtwinning as a cause (all our crystals have some degree of twinning). If the twin-plane alignment is predominantly along one of the body

diagonals, one might suppose that 'guided motion' of the vortices constrains them to diffuse at an angle, thereby generating the rotated ridge (Fig. 4a). But this cannot account for the change in sign of the rotation with field (guided motion is unaffected by field direction; Fig. 4b). As a further check, we turned the sample over and repeated the experiment with a positive field (Fig. 4c). If guided motion were responsible, the ridge rotation should now appear to have changed sign. This disagrees with our experiment (Fig. 2 shows that the rotation sense is unchanged when the sample is turned over from its orientation in Fig. 1). Similar arguments eliminate inhomogeneous distribution of pinning centres and variations of crystal thickness as candidates.

Because the rotation of the ridge changes sign with the field, it is natural to ask if the 'flux-flow' Hall effect, usually neglected when discussing vortex diffusion, is important here. When vortices depin and move freely in response to a current density J_c , their velocity v_L makes a small angle θ_H to $J_c \times \mathbf{B}$, instead of being parallel to it (Fig. 4d)⁸. This slight tilt introduces a 'longitudinal' velocity components $v_{\parallel}$ that is either parallel or antiparallel to J_c . (Because the transverse electric field $\mathbf{E}_H$ generated by $v_{\parallel}$ changes sign with $\mathbf{H}$, it is observed as a Hall signal.) Hall results in the low- T_c superconductor⁹ 2H-NbSe₂ are in good agreement with the theory of Nozières and Vinens⁸ but the Hall effect in the cuprates is less well understood^{10–12}.

In the usual picture of flux expulsion in a decreasing field, the outwardly diffusing vortices move along $J_c \times \mathbf{B}$, to produce a symmetric remanent profile at zero field. But if the Hall angle is large, the vortex velocity acquires a component $v_{\parallel}$ (parallel to J_c everywhere), which imparts a circulation to the outwardly diffusing vortex velocity (Fig. 4d). One might then argue that when the field reaches zero, the circulation produces a ridge that appears rotated. Moreover, it is easy to see that the sense of rotation reverses with field direction. As we shall show,

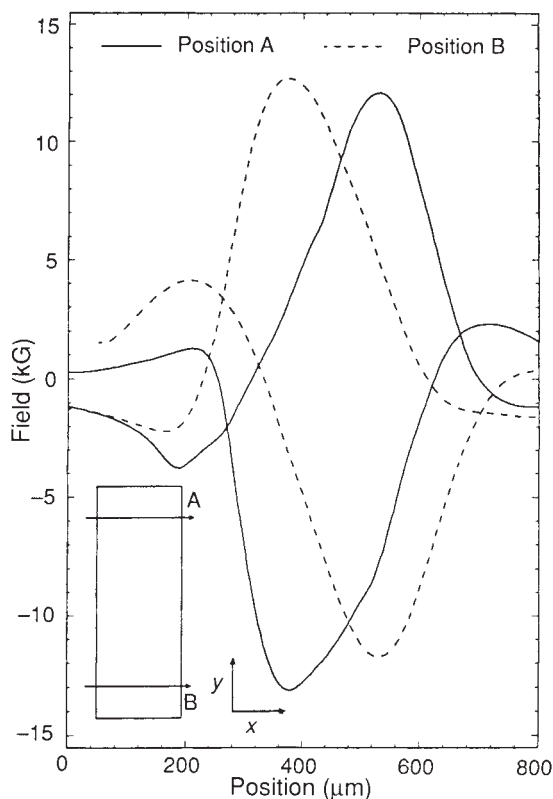


FIG. 1 Comparison of the remanence profiles at A (solid curves) created with positive and negative fields ($\mathbf{H} \parallel \pm \hat{z}$). The positive peak is displaced by 130 μm to the right of the negative peak. Similar scans at B show the opposite displacement (broken curves). The inset shows the scan positions to scale (scans are 140 μm from the edges). The crystal (sample 1) has an a – b face of $1,100 \times 400 \mu\text{m}^2$, and a thickness of 60 μm .

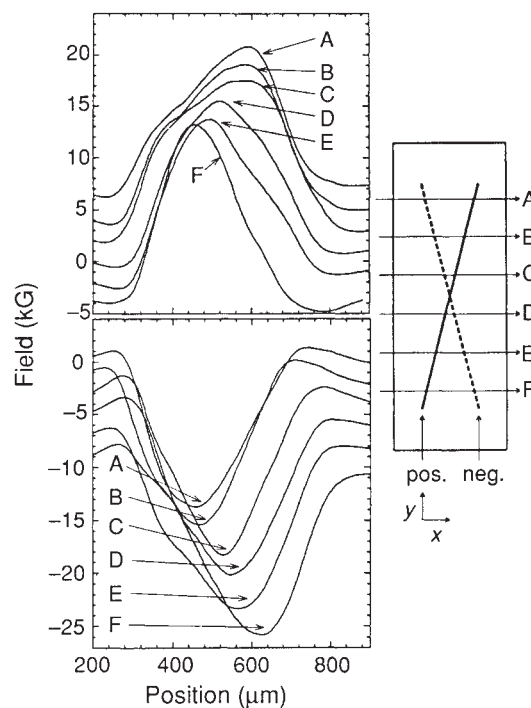


FIG. 2 Upper panel: scans of the positive-field remanence for different y , showing the shift of the peak to the left as we move from the top to the bottom half of the a – b face (A–F). Individual curves are offset by 2 kG for clarity. The scans are separated by 140 μm . Lower panel: negative-field scans along the same sections as in upper panel. The ridges for the two field orientations are shown to scale as solid and broken lines on the right. (The crystal is the same as in Fig. 1, but has been turned over by rotating about the y -axis.)

however, an explanation of this type runs into problems related to the magnitude and sign of the observed effect.

The observed rotation in Fig. 2 requires a large θ_H , whereas Hall measurements¹⁰⁻¹² in YBCO at high temperatures reveal a rather small θ_H (at 80 K, θ_H is $\sim 6 \times 10^{-3}$ rad T^{-1}). For the Hall effect to be relevant, we need to assume that θ_H increases markedly with decreasing T , to 50° or more in a field of 1 T. Interestingly, a simple extrapolation to 4.2 K of the unusual $1/T^2$ temperature dependence of θ_H in the normal state ($T > 100$ K), does produce a Hall angle of the right order¹³. In any case, a large θ_H at low temperatures cannot be precluded *a priori*. A large Hall angle would be unusual in a type II superconductor. The behaviour of vortices in the large- θ_H limit is relevant to unresolved issues in theories of vortex motion⁹.

A more serious problem arises from the sign of the rotation. Superconductors in which $v_{\parallel}$ is parallel (antiparallel) to $\mathbf{J}$ are said to show a positive (negative) flux-flow Hall effect. Measurements on YBCO show that its flux-flow Hall effect is predominantly positive, except for a small temperature range from 85 to 90 K, where it is negative¹⁰⁻¹². From 85 to 75 K (the lowest temperature for which Hall data are available), the Hall effect remains positive at all fields up to 15 T. A positive Hall effect implies that the circulation of the outwardly diffusing vortices has the same circulating sense as $\mathbf{J}_c$, implying that $\hat{\Omega}_{out} \parallel \mathbf{H}$. This disagrees with experiment (a negative Hall effect is needed to generate the clockwise ridge rotation in Fig. 4d). Thus, extrapolating the known behaviour of the flux-flow Hall angle from 74 to 4.2 K gives the wrong sign for Ω_{out} . This sign problem has an interesting twist. Let us apply the Hall model to the case when flux invades a sample that is initially free of vortices ($\mathbf{H}$ is increased from zero along $+\mathbf{z}$). Without the Hall effect, the flux front invades from all sides, leaving a deep trough in B

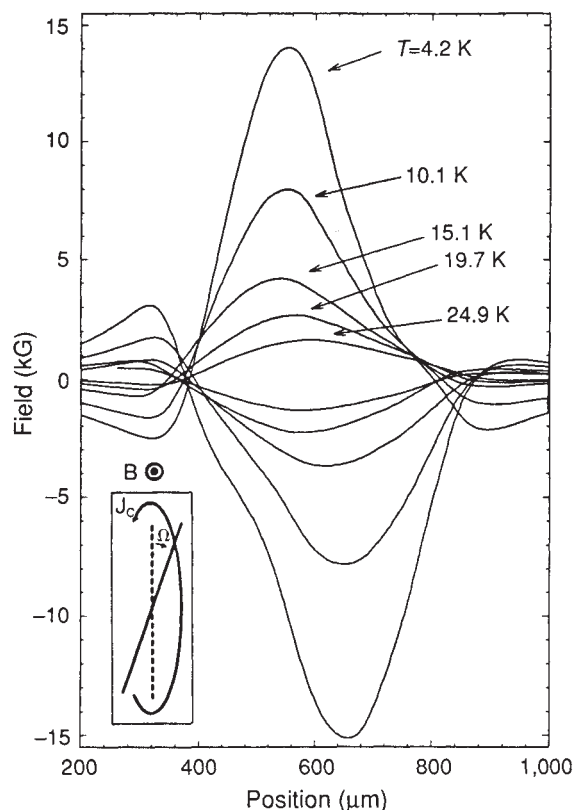


FIG. 3 Variation with temperature of the remanent profiles in sample 1 for the two field directions at $y = 150 \mu\text{m}$, measured from the upper edge. The asymmetry is not resolved above 25 K, but becomes increasingly prominent with decreasing temperature. The inset shows the direction of the critical state current and the rotation of the ridge for positive field.

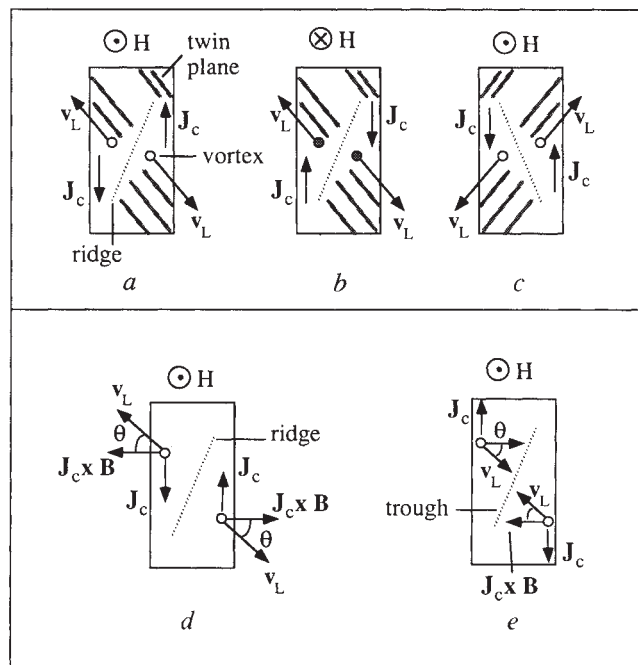


FIG. 4 Upper panel: vortex velocities during field sweep-down in 'guided motion' model. When remanence is set up in a positive (a) or negative (b) field, guided motion of vortices along twin boundaries (assumed to be aligned) may lead to a rotated ridge. Model predicts that the rotation of the ridge has the same sign (clockwise) for either field direction. If the crystal is turned over (c) the ridge should appear rotated in the anticlockwise direction. Both predictions conflict with experiment. Lower panel: vortex velocities during field sweep-down (d) and sweep-up (e) in the 'Hall-angle' model. During sweep down, outward flow velocity is tilted at angle θ_H to $\mathbf{J}_c \times \mathbf{B}$. A negative Hall angle is drawn in d to force agreement with observation. The remanence ridge is rotated clockwise. (If the field is negative, the ridge rotation is anticlockwise.) For inward flowing flux (e), however, the model predicts a clockwise rotation for positive field. This disagrees with experiment.

that is symmetric and parallel to a long side (an inverted roof). A large Hall angle imparts a circulation to the inward diffusion (Fig. 4e), resulting in a rotation of the trough (represented by the vector $\hat{\Omega}_{in}$). Because the circulation of $\mathbf{J}_c$ is now clockwise, and v_L points inwards, the Hall model predicts that the sign of the rotation is unchanged from the remanence experiment, $\hat{\Omega}_{in} \parallel -\mathbf{H}$. We have done this test and find the opposite result: $\hat{\Omega}_{in}$ is parallel to $\mathbf{H}$. For the Hall model to agree with experiment, we would need to associate outward diffusion of vortices with a negative Hall angle, and inward diffusion with a positive angle. These observations impose rather stringent tests, and it seems that a more sophisticated theory than the simple Hall-angle model is required. □

Received 4 May; accepted 7 July 1992.

- Yeshurun, Y. & Malozemoff, A. P. *Phys. Rev. Lett.* **60**, 2202-2205 (1988).
- Koch, R. *et al.* *Phys. Rev. Lett.* **63**, 1511-1514 (1989).
- Gammel, P. L., Schneemeyer, L. F. & Bishop, D. J. *Phys. Rev. Lett.* **66**, 953-956 (1991).
- Brandt, E. H. *Physica C* **185-189**, 270-275 (1991).
- Bean, C. P. *Rev. mod. Phys.* **36**, 31-39 (1964).
- Ong, N. P. *et al.* *Physica C* **153-155**, 1072-1077 (1988).
- Wang, Z. Z., Ong, N. P. & McGinn, J. *J. appl. Phys.* **65**, 2794-2798 (1989).
- Nozières, P. & Vinen, W. F. *Phil. Mag.* **14**, 667-688 (1966).
- Jing, T. W. & Ong, N. P. *Phys. Rev. B* **42**, 10781-10784 (1990).
- Iye, Y., Nakamura, S. & Tamegai, T. *Physica C* **154**, 616-619 (1989).
- Hagen, S. J., Lobb, C. J., Greene, R. L., Forrester, M. G. & Kang, J. H. *Phys. Rev. B* **41**, 11630-11633 (1990).
- Chien, T. R., Wang, Z. Z. & Ong, N. P. *Phys. Rev. Lett.* **66**, 3075-3078 (1991).
- Chien, T. R., Brawner, D. A., Wang, Z. Z. & Ong, N. P. *Phys. Rev. B* **43**, 6242-6245 (1991).

ACKNOWLEDGEMENTS. We thank P. W. Anderson, P. Nozières and T. V. Ramakrishnan for discussions, and the Office of Naval Research and the Seaver Institute for support.

Structure of stratospheric intrusions into the troposphere

C. Appenzeller & H. C. Davies

Institute for Atmospheric Physics, ETH, Höggerberg HPP, CH-8093 Zürich, Switzerland

FLOW phenomena that link the stratosphere and the troposphere are a key component of mid-latitude weather systems¹, and can influence substantially the distribution of atmospheric gases and aerosols^{2,3}. Little is known, however, about the detailed structure of these features. Here we combine satellite measurements with data from a weather prediction model to show that fine-scale structure can be resolved in intrusions of stratospheric air into the troposphere. We observe intrusions that develop into elongated (about 2,000 km) and slender (about 200 km) streamers which break up into a train of vortex-like disturbances while the streamer's tip splits and/or rolls up. The appearance of these structures might influence the daily development of weather systems, and will induce local, irreversible isentropic mixing of stratospheric air with the troposphere, possibly accounting for a substantial portion of the net stratosphere-to-troposphere exchange.

Major intrusions of stratospheric air down into the troposphere occur across the tropopause break, the sharp transition of the troposphere's depth from ~17 km in the tropics to ~9 km in polar regions. Typically an intrusion occurs in association with the development of a low-pressure system and a cold front at the ground, and a tongue of stratospheric air subsides isentropically towards the tropics to the west of the northeast-southwest orientated front. In a vertical section perpendicular to the front it forms a deep (~3–5 km) V-shaped fold^{3–5}. On the gently sloping (~1:100) isentropic surfaces the ~400-km-wide tongue evolves further either by curving cyclonically (anticlockwise in the Northern Hemisphere) to form a ~1,000-km-scale vortex above the surface low-pressure system^{6,7}, or alternatively by elongating anticyclonically to form a streamer^{1,8,9}.

In comparison with upper-troposphere air, stratospheric intrusions are rich in potential vorticity (PV) and ozone, deficient in water vapour and fluorocarbons, and differ in their aerosol content. In general, at least a part of these intrusions does not return unmodified into the main body of the stratosphere. Previous studies have shown that they can be either mixed with tropospheric air by small-scale turbulence acting at the lateral borders of the fold^{3,4,10}, or sequestered within the vortex located above the surface low-pressure system¹¹. Here in contrast we demonstrate that the streamer-variant of the intruded tongue can develop a structure intermediate in scale between small-scale turbulence on the fringe of the fold and large-scale vortex formation.

The evidence we present is in two forms: imagery of radiance measurements in the 5.7–7.1- μm water vapour band, and the contemporaneous potential vorticity distribution on an isentropic surface that slopes downwards from the pole and passes through the tropopause break in mid-latitudes. Within an intrusion, both water vapour and PV are conserved variables; the water vapour is merely a passive tracer but the PV is linked dynamically to the flow. The water vapour images are from Meteosat 4, a satellite of the European Space Agency (ESA), and provide a grey-scale spatial depiction of the radiance field with a pixel resolution of $\sim(8 \times 5)$ km² in mid-latitudes¹². In cloud-free conditions the radiance values constitute a weighted measure of the net water vapour content between ~3 and 9 km, and for standard conditions the contribution to the total radiance peaks between 6 and 7 km (refs 13, 14). An intrusion of dry stratospheric air slanting several kilometres into the troposphere will modify the outgoing radiance so that a part of the fold's entrance will appear as a 'dark' dry-band^{9,13}. A vertically orien-

tated intrusion ~4 km deep will reduce the precipitable water content of the upper troposphere by more than half, and this can yield an increase in the radiance by a factor of a third¹⁵. The second form of display, the PV distribution on an isentropic surface, is calculated from the pseudo-analysis fields of the operational weather prediction model (T213) of the European Centre for Medium-range Weather Forecasts (ECMWF). These fields are a synthesis of current observations using a statistical interpolation scheme with the model's own six-hour forecast providing the background state. The model itself has a horizontal and vertical resolution of ~60 km and ~1 km respectively, at tropopause levels in mid-latitudes¹⁶: this is finer-scaled than the upper troposphere observations available over Europe. Hence the smaller-scale structures in the PV pattern are artefacts of the model which are not at variance with the observations. In this display on an isentropic surface that transits into the troposphere in mid-latitudes, an intrusion is indicated by a tongue of high PV extending towards the Equator from the PV-rich stratosphere¹. For an intrusion event the two forms of display are complementary: the dark band in the water vapour image is a vertically averaged, but horizontally fine-scaled, depiction of the streamer, and the contiguous tongue on the isentropic surface is a coarse resolution depiction of the dynamically important PV tracer.

Features of the synoptic climatology of central and southern Europe^{17–20} suggest that it is a favourable region for streamer-

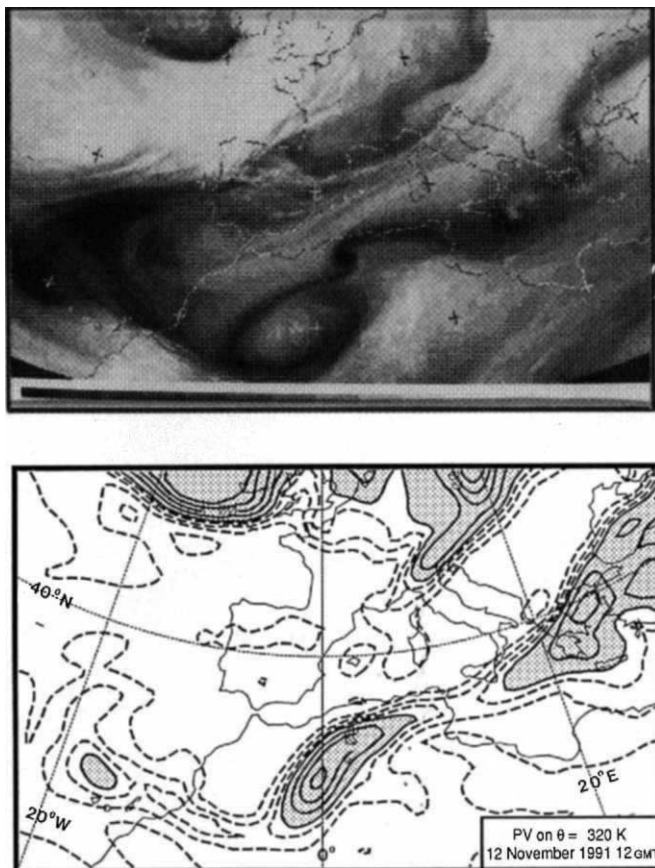


FIG. 1 Two depictions of the atmospheric structure over southern Europe at 12:00 GMT on 12 November 1991. The upper panel is a grey-scale display of radiance measurements from the 5.7–7.1- μm water vapour band. Dark regions denote large radiance values (and anomalously low moisture content in the mid-to-upper troposphere). The lower panel is the potential vorticity (PV) distribution on the 320 K isentropic surface. Contours are dashed for PV < 2 (with spacing of 0.5 units) and continuous for PV > 2 (with 1.0 spacing). PV values greater than 2 PV units are shaded and signify air of essentially stratospheric origin.

type intrusions. Here we provide two illustrative examples of such events. In the first case a time sequence of PV charts (not shown) indicated that the initial intrusion progressed southward for several days, curved anticyclonically and penetrated downwards $\sim 3\text{--}5\text{ km}$ into the troposphere. In its mature phase it formed a streamer stretching south over Europe and its southern extremity extended westward across southern Europe to the Mediterranean and on over North Africa and the eastern Atlantic (Fig. 1). The satellite image shows a slender, highly structured streamer with, as we scan westward along the band, wave disturbances giving way to vortex-like features and culminating in a rotated λ -shaped structure at the western end. The PV streamer on the isentropic surface has a more smeared appearance, but the distinct centres over Algeria and the southern Aegean correspond to almost coincident vortex-like structures in the radiance fields. A local PV centre is associated with a cyclonic flow component and a wind signature of this form was detectable beneath the centre over Algeria at heights of $\sim 3\text{ km}$. The displays are mutually supportive: the water vapour image indicates the streamer's fine-scaled structure and the PV distribution points to its dynamical nature. Taken together, they confirm the stratospheric origin of the streamer and indicate that it is breaking up into a train of vortices.

For the second case (Fig. 2), the satellite image reveals a richly structured anchor-shaped remnant of a matured streamer extending over Europe, and another maturing streamer, aligned

northwest-southwest, approaching the continent. These streamers also possess wave and vortex-like structures, for example the vortex roll-up above the Baltic and the growing wave and vortex features of the Atlantic streamer. Corresponding features, albeit with less fine-scale structure and slight spatial shift, are evident on the contemporaneous isentropic chart. The anchor-shaped remnant mentioned above is embedded in a region of enhanced PV, and a large, deep and cloud-free high-pressure system dominates the surface pattern. In contrast, the Atlantic streamer and its main vortex feature are (and subsequently remained) aligned with a growing wave on a surface cold front. This kind of development is consistent with the expectation that an upper-level PV vortex should, in the presence of a surface front, instigate the development of new weather systems^{1,22}.

The initial intrusion of a streamer can be viewed as an integral part of the day-to-day development of synoptic scale ($\sim 3,000\text{ km}$ wide) high- and low-pressure systems in mid-latitudes¹. Ambient mean anticyclonic flow then favours the streamer variant of the intrusion's evolution^{8,21,22}. The emerging streamer is subject to both deformation by the large-scale flow and the nonlinear dynamics of the PV band itself. One response to these competing effects is the roll-up and/or splitting of the rearward tip of the streamer, and another is the breakup of the streamer's main stem into waves and vortices. The breakup is consistent with the occurrence of an instability associated with the streamer's band of enhanced potential vorticity²³⁻²⁵. This instability yields a train of vortex disturbances with a growth rate comparable to the lateral PV gradient within the band and a wavelength ~ 5 times its half-width. This is analogous to classical shear-zone instability. Successive stages in the streamer's subsequent development resemble, and are dynamically related to, Rossby wave-breaking which erodes the polar vortex at higher elevations in the stratosphere^{26,27}, the formation of small-scale cut-off lows in the upper troposphere^{18,27}, the ubiquitous train of frontal waves at the Earth's surface^{23,24} and the typical pattern of cold and warm fronts within mid-latitude low-pressure systems.

The streamers in Figs 1 and 2 are typical of events that occurred with a frequency of $\sim 5\text{--}7$ days over western Europe during the winter of 1991-92. The frequency with which streamers occurred, coupled with their intensity, slenderness and breakup, has interesting ramifications. The existence of the streamer and its breakup into distinct dynamical entities indicates that, on the $100\text{--}1,000\text{-km}$ scale, the upper-level flow pattern can have a structure nearly as rich as that at the Earth's surface. Moreover, because the distribution of potential vorticity is intrinsically linked to the flow¹, the streamer's breakup changes its *in situ* and far-field flow signature, both becoming more vortex-like. In turn the vortices can instigate or modify the development of weather systems, one example being the frontal wave beneath the Atlantic streamer. To capture such development, the finer-grained features of streamers must be adequately represented in the initial conditions of a forecast, but their scale is at or beneath the resolution of many observational networks and prediction models, which might have an adverse effect on the forecasts.

A streamer's breakup into vortices also indicates local irreversible mixing on isentropic surfaces. This instability-induced mixing operates on the $\sim 100\text{-km}$ scale, involves the entire streamer rather than merely its bounding surface and ensures that the intruded air does not return to the stratosphere. The accompanying exchange of chemical constituents and aerosols can induce chemical and radiative effects^{3,10,28,29}, and there is also a transport of aerosols into domains where their residence times are very different². These processes are important in climate studies. The scale and frequency of the breakup of streamers is such that they can accomplish a substantial portion of the net global stratosphere-to-troposphere exchange. Moreover, the preferred development of streamers in certain longitudinal sectors implies that the time averages of the chemical and radiative changes

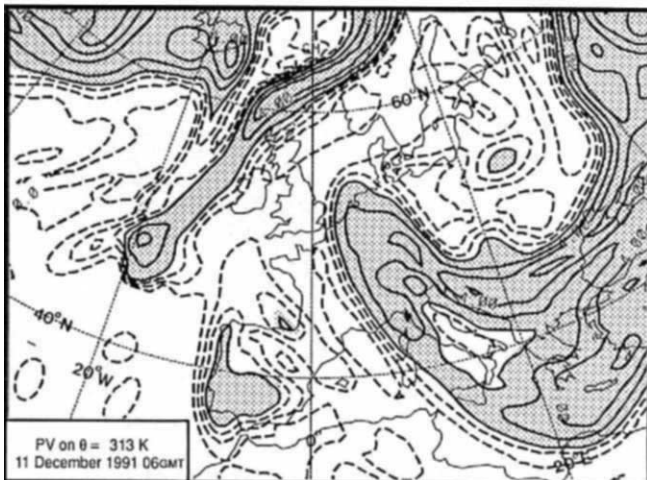
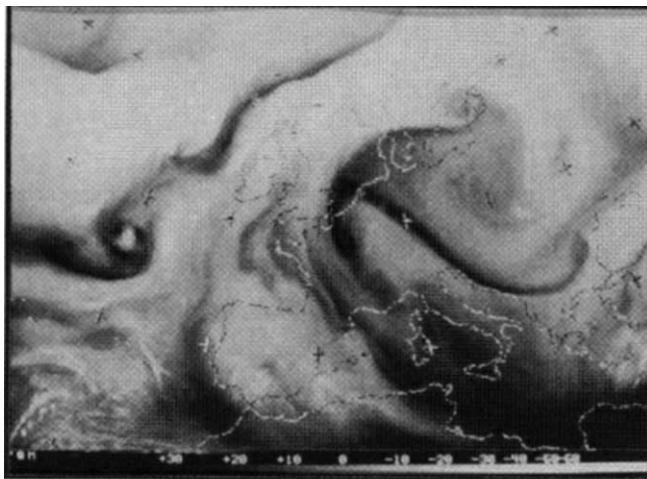


FIG. 2 As for Fig. 1, but for 06:00 GMT on 11 December 1991. Note that the PV distribution is now for the 313 K surface and that the geographical domain is slightly different.

will also show a longitudinal variation and this has implications for the interpretation of the recently documented longitudinal variations in ozone trends³⁰. □

Received 10 April; accepted 3 July 1992.

- Hoskins, B. J., McIntyre, M. E. & Robertson, A. W. *Q. J. R. met. Soc.* **111**, 877–946 (1985).
- Danielsen, E. F. *J. Atmos. Sci.* **25**, 502–518 (1968).
- Shapiro, M. A. *J. Atmos. Sci.* **37**, 994–1004 (1980).
- Danielsen, E. F. *et al. J. geophys. Res.* **92**, 2103–2111 (1987).
- Keyser, D. & Shapiro, M. A. *Mon. Weath. Rev.* **114**, 452–499 (1986).
- Whitaker, J., Uccellini, L. W. & Brill, K. F. *Mon. Weath. Rev.* **116**, 2337–2365 (1988).
- Browning, K. A. *Extratropical Cyclones: The Erik Palmén Memorial Volume*, (eds Holopainen, E. O. & Newton, C. W.) 129–154 (American Meteorological Society 1990).
- Hoskins, B. J., *Extratropical Cyclones: The Erik Palmén Memorial Volume*, (eds Holopainen, E. O. & Newton, C. W.) 64–90 (American Meteorological Society 1990).
- Müller, B. M. & Fuelberg, H. E. *Mon. Weather Rev.* **118**, 705–722 (1990).
- Danielsen, E. F. *et al. J. geophys. Res.* **96**, 17433–17452 (1991).
- Bamber, D. J. *et al. Atmos. Envir.* **18**, 1759–1766 (1984).
- Honvault, C. *Weather Satellites: Systems, Data, and Environmental Applications* (eds Rao, P. K. *et al.*) 69–76 (American Meteorological Society, 1990).
- Ramond, D., Corbin, H., Desbois, M., Szejwach, G. & Waldteufel, P. *Mon. Weath. Rev.* **109**, 2164–2176 (1981).
- Fischer, H., Eigenwillig, N. & Müller, H. *J. appl. Met.* **20**, 1344–1352 (1981).
- Poc, M. M., Rouleau, M., Scott, N. A. & Chedin, A. *J. appl. Met.* **19**, 868–876 (1980).
- Simmons, A. J. *Numerical Methods in Atmospheric Models* 281–324, (European Centre for Medium-range Weather Forecast, 1991).
- Lau, N.-C., White, G. W. & Jenne, R. L. *Circulation Statistics for the Extratropical Northern Hemisphere based on NMC Analyses* (National Center for Atmospheric Research, Boulder, Colorado, 1981).
- Holopainen, E. O. & Rontu, L. *Tellus* **33**, 351–359 (1981).
- Liasat, M. C. & Puigcerver, M. *Met. atmos. Phys.* **42**, 171–178 (1990).
- Price, J. D. & Vaughan, G. *Ann. geophys.* **10**, 96–102 (1992).
- Davies, H. C., Schär, C. & Wernli, H. *J. Atmos. Sci.* **48**, 1666–1689 (1991).
- Thorncroft, C. & Hoskins, B. J. *J. Atmos. Sci.* **47**, 2317–2336 (1990).
- Schär, Ch. & Davies, H. C. *J. Atmos. Sci.* **47**, 929–950 (1990).
- Joly, A. & Thorpe, A. J. *Q. J. R. met. Soc.* **116**, 525–560 (1990).
- Dritschel, D. G., Haynes, P. H., Juckes, M. N. & Shepherd, T. G. *J. Fluid Mech.* **230**, 647–665 (1991).
- McIntyre, M. E. & Palmer, T. N. *J. Atmos. terrest. Phys.* **46**, 825–849 (1984).
- Juckes, M. N. & McIntyre, M. E. *Nature* **328**, 590–596 (1987).
- Scientific Assessment of Stratospheric Ozone: 1989*, WMO Ozone Rep. No. 20, (WMO, 1990).
- Kelly, K. K., Tuck, A. F. & Davies, T. *Nature* **353**, 244–247 (1991).
- Stolarski, R. *et al. Science* **256**, 342–349 (1992).

Interference imaging of daily growth bands in massive corals

M. J. Risk* & T. H. Pearce†

* Department of Geology, McMaster University, Hamilton, Ontario, Canada L8S 4M1

† Department of Geological Sciences, Queen's University, Kingston, Ontario, Canada K7L 4V1

CORAL skeletons have been used to monitor a wide range of environmental parameters, including water temperature, insolation, siltation, river runoff and pollutant concentrations^{1–9}. In most cases, interpretation of the coral record relies on a determination of the growth rate. Radiographic techniques^{9–13}, which have been verified by a number of approaches^{8,14,15}, are most commonly used for this purpose, but are restricted to the determination of yearly average growth rates. Interpolation procedures have been used to extract growth rates on shorter timescales, but their accuracy depends on various assumptions about relative growth rates of skeletal bands with different densities^{9,16}. Here we report that reflectance optical microscopy can be used to measure daily growth rates of massive corals. This method should greatly expand the potential use of corals as environmental indicators.

Corals show pronounced daily growth rhythms, so it might be expected that some record of these rhythms would be preserved in the skeletons^{14,17}. Growth bands identified as daily or circadian have previously been reported in the epitheca of corals^{14,18–20}, but epithecal banding patterns are of limited use in environmental studies of coral growth. Images of sections through the skeletons of modern massive scleractinian corals have been published that contain possible monthly or daily growth increments^{20,21}. These images, however, have been mainly scanning electron micrographs, X-ray radiographs and transmitted-light photomicrographs.

Transmitted-light microscopy is not an ideal technique for examining structures such as daily growth bands; the images can be ambiguous because object space in transmitted-light microscopy is a volume rather than a surface. Details within this volume are neither easily resolved to the eye nor clearly preserved on a film plane. The clearest transmitted-light images of growth bands published so far are not consistently detailed enough for extensive measurement of growth rates on timescales of less than a year.

To obtain the clearest possible images of growth banding, we look at an etched surface and use reflected-light Nomarski differential interference contrast to image the surface. Reflected-light Nomarski imaging²² is an established technique recently used to study growth zones in magmatic plagioclase and the fine detail of volcanic rocks^{23–25}. The technique greatly increases the apparent relief of the microtopography of the surface, revealing features that would not otherwise be visible. Because we use a polished flat surface in reflected light, fine textural features may be seen without the confusion of detail present in transmitted light scenes. The resolution of the Nomarski technique is that of optical light microscopy (0.4 μm). The technique is well suited for imaging daily growth features, which are estimated to be 10–20 μm thick for most common massive corals.

We cut a carefully oriented vertical section through a coral specimen, then mount the slab in blue-dyed epoxy. We then grind the specimen flat and use tin oxide paste to obtain an optically polished surface. In our experience, tin oxide gives a better final polish than diamond paste, and a fine polished surface is essential for satisfactory results. The polished sample is etched in 2% acetic acid for up to 10 s. The etch is stopped with a bath of concentrated sodium carbonate solution. In some cases, the contrast of the image was improved by carbon-coating the specimen.

The etching process produces a microtopographic relief in the coral skeleton of $\sim 0.5 \mu\text{m}$ on the specimen surface. After etching, the sample retains its polished appearance; the slight microtopography is visible in reflected light only if interference microscopy is used. The resulting images were photographed, and measurements made on positive prints, directly on negatives and, in some cases, from long records improved by computerized image enhancement.

We studied sections of different corals including *Siderastrea siderea* from Costa Rica and Jamaica, *Diploria strigosa* from Tobago, *Porites lobata* from the Maldives and the East Pacific, and *Montastrea annularis* from Florida. All of our vertically oriented sections were cut from coral colonies whose growth rate had previously been established by radiography, staining or a combination of both techniques. All of the corals studied had fine growth bands 'microbands'; Fig. 1, and in every case the number of microbands per radiographic or stain-determined 'year' was very close to 365. The corals came from a wide variety of tidal regimes, but we saw no evidence of tidal control on the microbanding. The microbands are produced by the etched edges on bundles of trabeculae that sweep around and lie parallel to skeletal structures such as septa and corallite walls. In some instances an individual microband can be traced all the way across the section.

A sample of *S. siderea* from an especially polluted area in Discovery Bay, Jamaica, was studied and images of microbanding recorded in digitized form. Overlapping prints (a total of 19) were enhanced using the 'Image' program on a Macintosh IIx computer. The resulting photomosaic is 2.7 m long, contains 1,012 microbands and represents just under 3 years' growth. The growth rate of this coral had been previously established by X-ray radiography. We counted microbands on a continuous traverse up the central part of the section. The number of microbands counted per radiographic 'year' was 366. The inherent precision of radiographic banding is probably $\sim \pm 5\%$ for any given year, but the errors are not additive from year to year.



FIG. 1 Nomarski interference photomicrograph of an etched polished section of the scleractinian coral *Siderastrea siderea*, from the coral reef at Cahuita, on the east coast of Costa Rica. Microgrowth increments can be clearly seen, developed at the edge of a corallite. In areas where the microbands cannot be seen, the bundles of aragonite fibres are parallel to the plane of the section and show no relief on etching. Carbon-coated specimen, 5 s etch. Scale bar represents 50 μm .

The total possible error in a continuous radiographic record is therefore the error for one year, or ~ 3 weeks (depending on the area of collection and the quality of the image). From the agreement of the microband counts on the long mosaic with the radiographically determined growth rate, we conclude that the microbands are formed daily. A further test was provided by a section cut from a head of *Montastrea annularis* from Florida, which contained a 4-year growth record, verified by both radiography and Alizarin Red "S" staining. This coral also showed microgrowth banding whose scale is consistent with a diurnal origin. The microbands were thick enough (20–30 μm) for a bimodal nature (possibly day and night growth) to be clear in each diurnal cycle.

From the photomosaic of the Jamaican *Siderastrea*, we measured relative growth rates in the dense and less-dense skeletal bands of an annual growth couplet. The width of the microbands is in part a function of their orientation. Where the bundles of crystallites are at high angles to the plane of section, the bands will appear wide. We took care to measure the zones at their narrowest points. In the dense bands, the coral grew 6.9 μm per day (four replicate counts with a total of 114 microbands, standard deviation 0.24). In the less-dense skeletal zones,

which were much wider, the coral grew 9.3 μm per day (four replicate counts with a total of 137 microbands, standard deviation 0.34). This correlates well with this particular coral's slow growth rate of 3 mm per year, and implies that it grew $\sim 25\%$ faster when laying down the less-dense skeletal band. This also demonstrates, as has already been shown, that yearly growth couplets in massive corals are produced by changes in the skeletal 'meso-architecture'²⁶.

Having established that the microbands were daily, we then applied the technique to a known example of environmental change. The 1982–83 El Niño event in the East Pacific was the largest in recent history, and resulted in mass mortality of corals in the East Pacific²⁷; the event is recorded in the few surviving corals¹⁰. We studied microgrowth increments in one such coral, from the west coast of Costa Rica. In 1980, before the incursion of warm water represented by the El Niño event, the coral grew 22 μm per day, or ~ 8 mm per year. The record of growth during and immediately after the event is difficult to interpret, but the maximum rate that we were able to obtain was 9 μm per day. Two years after the El Niño, in 1985, the growth rate was 18 μm per day, or about 6.6 mm per year. The skeletal structure and chemistry of these particular corals is well known, including the growth rate before the El Niño event, the extreme decrease or cessation of growth during the event, and the growth depression for 2–3 years after²⁸. The growth rate estimates based on microbands agree with the radiographic estimates to within 0.1 mm.

To interpret fully the record of daily growth, we need a better understanding of the relationship between growth and trabecular orientation and spacing. Although all the corals we studied showed microgrowth banding, long continuous, unambiguous banding sequences were generally hard to find (the orientation of sections is critical in this respect).

The technique described here might greatly expand the ability of corals to serve as monitors of environmental change. It may now be possible to study the impact of short-lived phenomena, such as isolated pollutant spills and volcanic eruptions, on coral growth. Coupling of microband and geochemical microprobe analysis might even pinpoint the precise time of uptake of a given contaminant. In addition, it should be possible to determine the maximum allowable levels of contaminants on coral coasts by controlled experiments that last days instead of years. □

Received 1 June; accepted 1 July 1992.

1. Macintyre, I. G. & Smith, S. V. *Proc. 2nd Int. Coral Reef Symp.* **2**, 277–287 (Great Barrier Reef Committee, Brisbane, Australia, 1974).
2. Druffel, E. M. & Suess, H. E. *J. geophys. Res.* **88**, 1271–1280 (1983).
3. Buddemeier, R. W., Maragos, J. E. & Knutson, D. W. *J. exp. mar. Biol. Ecol.* **14**, 179–211 (1974).
4. Shinn, E. A. *Paleontology* **40**, 233–240 (1966).
5. Dodge, R. E. & Gilbert, T. R. *Mar. Biol.* **82**, 9–13 (1984).
6. Brown, B. E. & Howard, L. S. *Adv. mar. Biol.* **22**, 1–63 (1985).
7. Cortes, J. & Risk, M. J. *Bull. mar. Sci.* **36**, 339–356 (1985).
8. Isdale, P. *Nature* **310**, 578–579 (1984).
9. Carriquiry, J. D., Risk, M. J. & Schwarcz, H. P. *Paleosols* **3**, 359–364 (1988).
10. Hudson, J. H., Shinn, E. A., Halley, R. B. & Lidz, B. *Geology* **4**, 361–364 (1976).
11. Shen, G. T. & Boyle, E. A. *Chem. Geol.* **67**, 47–62 (1988).
12. Woodroffe, C. & McLean, R. *Nature* **344**, 531–534 (1990).
13. Dodge, R. E. & Vaisnys, J. R. *Nature* **258**, 706–708 (1975).
14. Barnes, D. J. *Proc. R. Soc. B* **182**, 331–350 (1972).
15. Risk, M. J., Pagani, S. E. & Elias, R. J. *Paleosols* **2**, 323–331 (1987).
16. Fairbanks, R. G. & Dodge, R. E. *Geochim. cosmochim. Acta* **43**, 1009–1020 (1979).
17. Chalker, B. E. & Taylor, D. L. *Proc. R. Soc. B* **201**, 179–189 (1978).
18. Ma, T. Y. H. *Proc. Imp. Acad. Tokyo* **10**, 353 (1934).
19. Wells, J. W. *Nature* **197**, 948–950 (1963).
20. Jell, J. S. *Proc. 2nd Int. Coral Reef Symp.* **2**, 301–320 (Great Barrier Reef Committee, Brisbane, Australia, 1974).
21. Barnes, D. J. & Lough, J. M. *J. exp. mar. Biol. Ecol.* **126**, 119–134 (1989).
22. Nomarski, G. & Weill, A. R. *Soc. France Miner. Crist. Bull.* **77**, 840–868 (1954).
23. Anderson, A. T. Jr. *Am. Miner.* **68**, 125–129 (1983).
24. Pearce, T. H. & Clark, A. H. *Geology* **17**, 757–759 (1989).
25. Pearce, T. H. & Kolisnik, A. M. *Earth Sci. Rev.* **29**, 9–26 (1990).
26. Barnes, D. J. & Devereux, M. J. *J. exp. mar. Biol. Ecol.* **121**, 37–54 (1988).
27. Glynn, P. W. *Envir. Cons.* **10**, 149–154 (1984).
28. Carriquiry, J. D. thesis, McMaster Univ. (1991).

ACKNOWLEDGEMENTS. We thank C. Kurucz and A. Kolisnik for laboratory assistance, C. Kurucz for data analysis, J. Whorwood for help with photography, and F. Cox and R. Halley for stained coral samples. This work was supported by the NSERC of Canada through Operating Grants to M.J.R. and T.H.P., and an Infrastructure Grant to support the Laser Laboratory at Queen's University.

An earthquake mechanism based on rapid sealing of faults

Michael L. Blanpied, David A. Lockner
& James D. Byerlee

United States Geological Survey, Mail Stop 977, Menlo Park,
California 94025, USA

RECENT seismological, heat flow and stress measurements in active fault zones such as the San Andreas have led to the suggestion^{1,2} that such zones can be relatively weak. One explanation for this may be the presence of overpressured fluids along the fault³⁻⁵, which would reduce the shear stress required for sliding by partially 'floating' the rock. Although several mechanisms have been proposed for overpressurizing fault fluids^{3,4,6,7}, we recall that 'pressure seals' are known to form in both sedimentary⁸ and igneous⁹ rocks by the redistribution of materials in solution; the formation of such a seal along the boundaries of a fault will prevent the communication of fluids between the porous, deforming fault zone and the surrounding country rock. Compaction of fault gouge, under hydrostatic loading and/or during shear, elevates pore pressure in the sealed fault and allows sliding at low shear stress. We report the results of laboratory sliding experiments on granite, which demonstrate that the sliding resistance of faults can be significantly decreased by sealing and compaction. The weakening that results from shear-induced compaction can be rapid, and may provide an instability mechanism for earthquakes.

We present the results from two experiments on granite 'faults' slid at hydrothermal conditions, that is, elevated temperature (T) and H_2O fluid pressure (P_p). Sawcut cylinders of granite (Fig. 1, insets) contained a layer of 'gouge' (fine granite powder, maximum particle size 90 μm). A bore hole drilled in the upper forcing block provided access for pore fluid. Experimental conditions were: confining pressure 400 MPa, $P_p = 100$ MPa, $T = 600$ °C. After resting 2 to 4 hours at these conditions, the samples were shortened at an axial rate of 8.66×10^{-5} mm s⁻¹. The coefficient of friction was calculated as the ratio of shear stress (τ) to effective normal stress ($\bar{\sigma}_n = \sigma_n - P_p$). We shall distinguish between apparent friction, μ_{app} , calculated using the nominal, imposed fluid pressure, and true friction, μ_{true} , calculated using the presumed fluid pressure in the fault itself, which may differ from that in the bore hole, as explained below.

Our first sample showed anomalously low strength (Fig. 1a). Elastic loading to a peak shear stress of 170 MPa ($\mu_{app} = 0.42$) was followed by a rapid, aseismic decay in strength to the remarkably low level of 76 MPa ($\mu_{app} = 0.22$). From independent tests at these conditions¹⁰ we expect μ to be in the range 0.46 to 0.52, more than twice as high as the observed value. This suggests that pore pressure in the sawcut was higher than the externally controlled 100 MPa. To test this hypothesis, we imposed large steps in P_p . These steps did not alter the shear stress, showing that there was no fluid communication between the bore hole and the fault surface. We conclude that the low strength resulted from elevated fluid pressure trapped within the deforming gouge layer. Further tests have shown that the low peak and residual stresses are reproducible, at the imposed pressures and strain rate, at $T = 550$ and 600 °C.

Given the sample geometry (Fig. 1a), the pressure seal probably formed within the intact granite of the forcing block ('country rock') that separated the gouge from the bore hole. To test this, we did a second experiment on a sample with a bore hole drilled through to the fault surface (drained). The results were in marked contrast to the previous experiment (Fig. 1b). On loading, a higher shear-stress peak was followed by a slow decay to the expected shear stress of 180 MPa ($\mu_{app} = 0.47$). Steps in P_p produced a large-amplitude response, showing that the pore fluid was able to communicate quickly with the fault.

Comparison of these two experiments demonstrates that the pressure seal was located within the granite country rock.

Presumably, the intact granite becomes sealed by the redistribution of materials in solution. While the granite is resting at conditions of high T and P_p , as well as during loading, solution-transport and precipitation of dissolved material seal the fluid pathways in the granite¹¹⁻¹³. Once the seal is in place, further compaction of the gouge causes P_p to rise. Compaction probably results from two distinct processes. First, the application of hydrothermal conditions initiates time-dependent compaction by stress-controlled dissolution-precipitation creep. Hydrothermal isostatic pressing experiments on sandstones¹⁴ and quartz powders^{15,16} show considerable compaction in a matter of hours. Second, the onset of simple shear within the underconsolidated gouge layer causes the porous structure to compact by the rearrangement and fracturing of particles^{17,18}. In our self-sealed experiments the onset of slip near to peak stress is followed by rapid weakening. We reason that slip causes compaction, rapidly raising the fluid pressure. This rise in P_p lowers the effective stress, resulting in rapid slip and loss of strength. By this argument, the difference between peak and residual strength is due mainly to the abrupt rise in P_p . An additional effect, contributing to rapid loss of strength following the onset of shearing, would be cementation or embrittlement of the gouge because of the solution-transport processes just described¹⁹.

The exact timing of seal formation is not known because we do not directly monitor fluid pressure within the gouge layer. The seal does not form instantly: in similar experiments published earlier¹⁰ we allowed a rest of only 1 hour and shortened the samples 10 times faster. No evidence for sealing or overpressure was seen in those tests. It seems likely, however, that P_p had begun to rise by the time the peak strength was reached (Fig. 1a), because this peak was lower than in the drained experiment (Fig. 1b). The ratio of peak strengths cannot be used to calculate P_p before slip begins because partial cementation of the gouge may have contributed to the magnitudes of the peaks²⁰.

We can calculate the pressure of trapped water needed to cause the low strength measured in the first experiment. The gouge is assumed to shear with $\mu_{true} = 0.46$ to 0.52. Then $\mu_{app} = 0.22$ in Fig. 1a requires P_p in the gouge layer to be 280–300 MPa. Because little or no water was able to pass through the sealed granite, this elevated pressure did not change even though P_p in the blind bore hole was stepped in the range 0 to 100 MPa. From the compressibility of water (ref. 21), we estimate that raising P_p from 100 to ~290 MPa in the sealed fault requires a loss of roughly 45% of the pore space. This figure seems to be consistent with data from the compaction studies cited above. We can also place a rough upper bound on the permeability of the sealed granite, using the duration of sliding at low stress (15 hours in one experiment), and reasoning that a significant loss of water from the fault would have caused a noticeable increase in shear strength. Given the sensitivity of our equipment, we estimate that the permeability of the granite was no greater than 10^{-23} m² (0.01 nanodarcies). (For comparison, the usual permeability of Westerly granite at an effective pressure of 300 MPa is $\sim 10^{-20}$ m² (10 nanodarcies; ref. 22).) This very small value is an upper bound: the permeability may well be much lower.

It is important to note that our experimental fault remained sealed and overpressured despite active shearing. This is possible because the granite making up the seal was not itself being deformed.

Although our experiments were done at 600 °C on granite, natural pressure seals can form at much lower temperatures, given sufficient time. Flow-through experiments on several crystalline rocks¹¹ have shown rapid sealing (hours to days) at temperatures of <250 °C. Laboratory data on the rate of hydrothermal crack healing in quartz^{13,23} indicate that the lifetime of crack-related permeability may be geologically short (years to

decades) at temperatures as low as 200 °C (ref. 24). In natural hydrothermal fields, seals consisting of precipitated quartz are commonly found near $T = 320$ °C (ref. 9), and carbonate seals capping oil and gas source rocks cluster around 95 °C (ref. 25). The years to decades that are typical interseismic time intervals should allow ample time to erect pressure seals, even at temperatures much lower than in our 600 °C experiments. In fact, there are many observations of natural faults in several rock types, suggesting that the generation and destruction of pressure seals are linked to cycles in the pressure of pore fluids^{26,27}. Given these observations, it seems likely that faults are bounded by sealed rocks well up into the seismogenic upper crust (typically the uppermost ~15 km).

We propose that compaction of porous fault-zone materials (gouge, breccia) causes fluid pressures to become elevated, lowering the fault strength. Others have proposed scenarios in which low fault strength results from high pore pressure. These models invoke either the expansion of pore fluids from coseismic shear heating³, or a mantle-derived source of fluids to maintain overpressure within a 'leaky' fault⁵, or require that the

fluid pressure gradient from the fault to the country rock should lie near a low 'threshold' level⁴. Here, we postulate that a fault may remain sealed indefinitely once pressure seals are in place, even though the fault may experience continuous creep or earthquakes. The shearing stress depends on the degree of overpressure, which in turn depends on the porosity and compaction rate within the fault. Our model of a sealed, shearing fault predicts low permeability across the fault, but finite permeability within the fault plane; this situation is required by the fault model of Rice. In the case of sub-vertical faults, in-plane permeability may allow the escape of fluids to the Earth's surface despite the sealing of the country rock. This phenomenon is inferred to occur at shallow levels within accretionary prisms, where fluids escape from the compacting, dewatering sediment wedge along actively deforming faults²⁸. For deeper faults, fluid escape may be prevented by impermeable fillings of clays⁷.

An implication of this model is that compaction will proceed, and P_p will rise, until fault strength is overcome, or until P_p exceeds the least principal stress, causing hydrofracture. Therefore, sliding (or hydrofracture) will eventually occur even for arbitrarily low differential stress. If pressure seals are widespread, large portions of the Earth's crust may be in a state of incipient failure.

Our results also suggest a novel instability mechanism for earthquakes. In our self-sealed experiments, the drop from peak to residual stress was rapid (250 MPa mm⁻¹) but did not constitute an inertial instability in our stiff testing machine. In the Earth, which has much greater compliance²⁹, a similarly abrupt weakening might lead to unstable slip. We propose that shear-induced compaction in a sealed, underconsolidated fault could cause an earthquake. We envisage an increment of slip causing compaction within the saturated fault zone; compaction causes pore pressure to rise, lowering the effective stress and initiating the instability. The stress drop and residual stress depend on the initial fluid pressure and the amount of compaction, and may differ greatly from our experiment. Pre-seismic slip can raise pore pressure only if the fluid is trapped within the fault zone. It is not necessary for permeability to be zero; it must only be low enough to prevent sufficient fluid to escape²⁴, and therefore depends also on the rates of pre-seismic creep and compaction and on the volume of compacting material. In a fully isolated fault, the creep rate may be arbitrarily low.

Our earthquake model does not predict in detail co-seismic behaviour, which may involve complicating factors such as shear heating³. But we note that for repeated earthquakes to be initiated by this mechanism in the same location on a fault, there must exist a means of re-establishing porosity. At least two possibilities exist. First, the collapse of porosity may involve only a portion of the available porous materials. Second, porosity may be re-established by damage accompanying seismic slip, so that the same fault materials could repeatedly undergo shear-induced compaction. Many exhumed faults show evidence for coseismic dilatancy^{27,30,31}.

Earthquakes associated with subducted lithospheric slabs occur to several hundred kilometres depth³². It has been argued that 'intermediate-depth' earthquakes (50–300 km) cannot result from frictional sliding because the normal stress across slip planes would require unreasonably high differential stresses. It is possible, however, that high fluid pressures are trapped in the descending slab, and that low effective stress allows frictional sliding despite the high confining pressure. Subduction of the ocean floor carries porous, water-laden sediments, plus hydrated oceanic upper mantle, into the Benioff zone^{28,33}. At the inferred temperatures of 250 to 600 °C (ref. 33), sealing by solution-transport should be rapid enough to trap excess fluids generated by compaction and dewatering reactions. It seems possible that the earthquakes observed to occur down to tens or even hundreds of kilometres result from the proposed mechanism of shear-induced compaction described above.

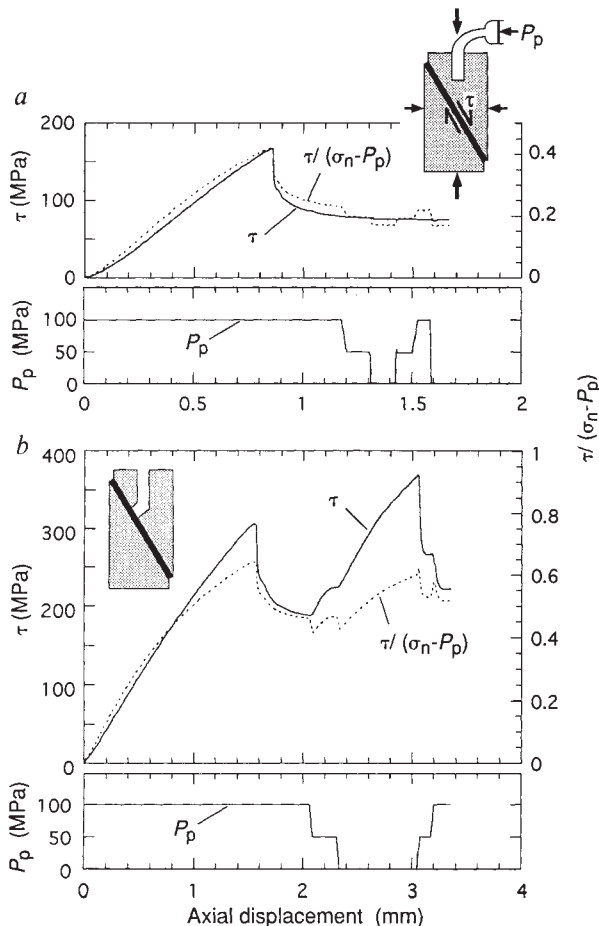


FIG. 1 Shear stress (τ), apparent friction ($\mu_{app} = \tau/(\sigma_n - P_p)$) and imposed H₂O pore pressure (P_p) in sliding tests on granite at 600 °C. Cross-section sketches of cylindrical granite samples (19 mm diameter by 38 mm long) show principal stress directions and sense of sliding. Sawcut 'fault' at 30° to the loading axis contains a 0.5-mm layer of granite powder. Bore hole in the upper forcing block admits H₂O. (For more on the method, see ref. 37.) a, Sample with blind bore hole shows elastic loading to a low peak stress. Onset of shear is accompanied by rapid decay of strength to a remarkably low level. Shear stress does not respond to large, imposed steps in P_p , showing that the sample has become sealed. Apparent friction responds because we calculate it assuming P_p in the fault is known. b, Sample with hole through to fault shows the expected higher peak and residual strengths. Shear stress responds strongly to steps in P_p , demonstrating fluid communication.

Hydrous phases may be carried even further down the subduction zone, perhaps to depths of 700 km or more³⁴. At these depths, however, water released through prograde reactions will be consumed in partial melts. Therefore, our mechanism for

producing earthquakes does not seem to explain the seismic maximum in the 400–660 km depth interval ('deep' earthquakes³²). These deepest events seem to require other processes^{34–36}. □

Received 11 March; accepted 18 June 1992.

1. Zoback, M. D. *et al. Science* **238**, 1105–1111 (1987).
2. Hickman, S. H. in *U.S. Nat. Rep. int. Un. Geodesy Geophys. 1987–1990, Rev. Geophys. Suppl.*, 759–775 (1991).
3. Lachenbruch, A. H. *J. geophys. Res.* **85**, 6097–6112 (1980).
4. Byerlee, J. *Geophys. Res. Lett.* **17**, 2109–2112 (1990).
5. Rice, J. R. in *Fault Mechanics and Transport Properties in Rocks (the Brace volume)* (ed. Evans, B. & Wong, T.-F.) 475–503 (Academic, London, 1992).
6. Sibson, R. H. *Nature* **243**, 66–68 (1973).
7. Fournier, R. O. *EOS* **71**, 1635 (1990).
8. Powley, D. E. *Earth Sci. Rev.* **29**, 215–226 (1990).
9. Fournier, R. O. *Geophys. Res. Lett.* **18**, 955–958 (1991).
10. Blanpied, M. L., Lockner, D. A. & Byerlee, J. D. *Geophys. Res. Lett.* **18**, 609–612 (1991).
11. Morrow, C. A., Moore, D. E. & Byerlee, J. D. *Mat. Res. Soc. Proc.* **44**, 467–473 (1985).
12. Summers, R., Winkler, K. & Byerlee, J. *J. geophys. Res.* **83**, 339–344 (1978).
13. Smith, D. L. & Evans, B. *J. geophys. Res.* **89**, 4125–4135 (1984).
14. Sprunt, E. S. & Nur, A. *Geophysics* **42**, 726–741 (1977).
15. Cox, S. F. & Paterson, M. S. *Geophys. Res. Lett.* **18**, 1401–1404 (1991).
16. Lockner, D. A. thesis, Massachusetts Inst. Technol. (1990).
17. Marone, C. & Scholz, C. H. *Geophys. Res. Lett.* **15**, 621–624 (1988).
18. Morrow, C. A. & Byerlee, J. D. *J. struct. Geol.* **11**, 815–825 (1989).
19. Angevine, C. L., Turcotte, D. L. & Furnish, M. D. *Tectonics* **1**, 151–160 (1982).

20. Fredrich, J. T. & Evans, B. *EOS* **72** (Fall mtg. Suppl.) 441 (1992).
21. Burnham, C. W., Holloway, J. R. & Davis, N. F. *Thermodynamic Properties of Water to 1,000 °C and 10,000 Bars* (Geol. Soc. Am., Boulder, Colorado, 1969).
22. Brace, W. F., Walsh, J. B. & Frangos, W. T. *J. geophys. Res.* **73**, 2225–2236 (1968).
23. Brantley, S. L., Evans, B., Hickman, S. H. & Crerar, D. A. *Geology* **18**, 136–139 (1990).
24. Walder, J. & Nur, A. *J. geophys. Res.* **89**, 11539–11548 (1984).
25. Hunt, J. M. *Am. Assoc. Petrol. Geol. Bull.* **74**, 1–12 (1990).
26. Ramsey, J. G. *Nature* **284**, 135–139 (1980).
27. Sibson, R. H., Robert, F. & Poulsen, K. H. *Geology* **16**, 701–704 (1988).
28. Behrmann, J. H. *Earth planet. Sci. Lett.* **107**, 550–558 (1991).
29. Walsh, J. B. *J. geophys. Res.* **76**, 8597–8598 (1971).
30. Chester, F. M. & Logan, J. M. *Pure appl. Geophys.* **124**, 79–106 (1986).
31. Power, W. L. & Tullis, T. E. *J. struct. Geol.* **11**, 879–893 (1991).
32. Frohlich, C. A. *Rev. Earth planet. Sci.* **17**, 227–254 (1989).
33. Wannamaker, P. E., *et al. J. geophys. Res.* **94**, 14127–14144 (1989).
34. Meade, C. & Jeanloz, R. *Science* **252**, 68–72 (1991).
35. Kirby, S. H. *J. geophys. Res.* **92**, 13789–13800 (1987).
36. Burnley, P. C., Green, H. W. & Prior, D. J. *J. geophys. Res.* **96**, 425–443 (1991).
37. Lockner, D. A., Summers, R. & Byerlee, J. D. *Pure appl. Geophys.* **124**, 445–469 (1986).

ACKNOWLEDGEMENTS. We thank J. Dieterich, B. Evans, S. Hickman, A. McGarr, D. Moore, N. Sleep and T.-f. Wong for discussions and reviews. This work is funded by the USGS programs in Earthquake Hazards Reduction and Deep Continental Studies.

Action spectrum for DNA damage in alfalfa lowers predicted impact of ozone depletion

F. E. Quaife, B. M. Sutherland & J. C. Sutherland

Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA

DEPLETION of stratospheric ozone will increase the intensity of solar mid-ultraviolet (280–320 nm) radiation reaching the biosphere¹. Predictions of increases in biologically effective ultraviolet radiation require knowledge of both the solar spectral intensity and the wavelength-dependent sensitivity (action spectrum) for damaging the biological target². A generalized action spectrum for plant damage encompassing wavelengths from 280 to 313 nm^{3–5} has been widely used to predict the consequences of ozone depletion. Calculations⁶ based on this spectrum and new satellite measurements of atmospheric ozone suggest that plants will be among those organisms most severely affected. Here we report an absolute action spectrum for cyclobutyl pyrimidine dimer induction in DNA in intact alfalfa seedlings, which reveals damage by wavelengths as long as 365 nm. Calculations based on this new action spectrum predict significantly smaller increases in biologically effective ultraviolet radiation resulting from ozone depletion, particularly at high latitudes, than calculations based on either the generalized plant action spectrum or the action spectrum for damaging unshielded DNA.

DNA in higher plants is shielded from environmental ultraviolet radiation by upper cell layers and pigments, in contrast to DNA in solution or in unpigmented single cells. Thus the action spectrum for DNA damage in plants and other higher organisms depends both on the DNA absorption spectrum and on the attenuation by shielding layers and pigments. We measured the wavelength dependence of DNA damage formation in intact alfalfa seedlings using an alkaline agarose gel electrophoresis method^{7,8} adapted for use in higher plants⁹.

Table 1 shows the data for pyrimidine dimer production in the seedlings, obtained from the average slope of the dose-response lines at each waveband. The action spectrum (Fig. 1) has a maximum near 280 nm, in contrast to the 260-nm maximum for unshielded DNA². This shift in maximum sensitivity results from intra-organism shielding and attenuation^{10,11}. Dimer formation was detected at wavelengths as long as 365 nm, but no

dimers were detected at 405 nm. Note that for wavelengths less than about 310 nm, the unshielded DNA is markedly more sensitive to ultraviolet damage, differing by a factor of more than 100 at 280 nm. But for wavelengths greater than 310 nm, the spectra for DNA damage induction in alfalfa seedlings and in human skin *in situ*¹² are similar to those for unshielded DNA¹³. As these are absolute action spectra for dimer formation,

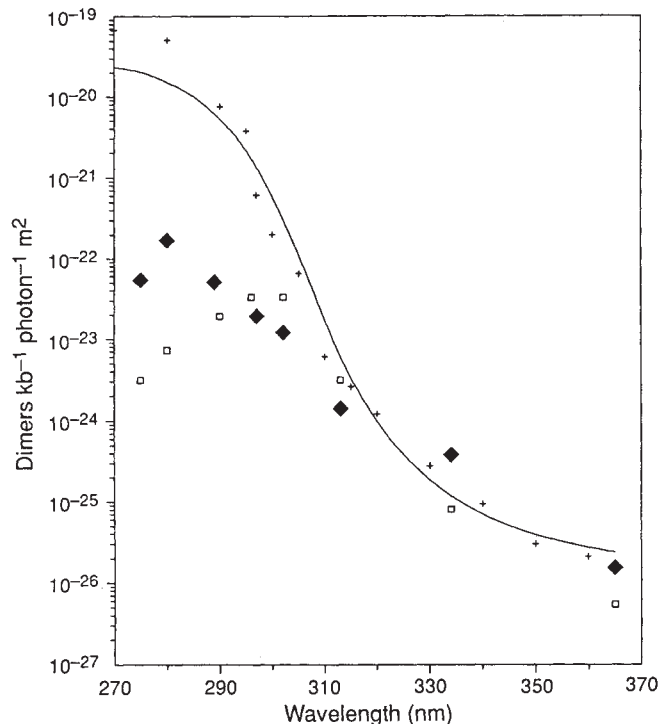


FIG. 1 Absolute action spectra for pyrimidine dimer induction in alfalfa seedlings (◆), T7 bacteriophage DNA¹³ (+) and human skin *in situ*¹² (□), expressed as the number of dimers formed per thousand bases of DNA per incident photon per square metre of irradiated surface⁸. The three absolute action spectra are similar for wavelengths greater than about 310 nm, but unshielded DNA is far more sensitive at shorter wavelengths. Setlow's DNA damage action spectrum² (—) was scaled to the T7 DNA data by a logarithmically weighted least-squares procedure applied at wavelengths from 280 to 320 nm.

TABLE 1 Action spectrum for pyrimidine dimer formation in DNA in alfalfa seedlings

Central wavelength (nm)	Spectral bandwidth (nm)	Filter	Dimer production cross-section ($\text{m}^2 \text{ photon}^{-1}$) $\times 10^{24}$
275	2	—	54 ± 4.5
280	1	—	170 ± 21
289	2	—	51 ± 4.6
297	2	—	19 ± 1.6
302	2	—	12 ± 3.2
313	2	310 nm cutoff (Mylar)	1.4 ± 0.24
334	4	310–400 nm bandpass + 320 nm cutoff	0.38 ± 0.08
365	4	310–400 nm bandpass + 360 nm cutoff	0.015 ± 0.0022

Alfalfa (*Medicago sativa* L.) seeds were germinated and grown for 6–7 days in a growth chamber at 20 °C with a 16-h photoperiod. Illumination was by cool white fluorescent bulbs filtered by UF4 plexiglass, which excludes wavelengths shorter than 400 nm. Seedlings were exposed (three independent experiments for each wavelength) to increasing doses of narrow-band ultraviolet radiation from a high-intensity monochromator¹⁵. For wavelengths above 310 nm, additional filters were used to ensure that shorter, more effective wavelengths were completely excluded. The incident intensities were determined with a Molecron PR200 pyroelectric radiometer. Immediately after irradiation cotyledons were minced, embedded in agarose plugs and digested with proteinase K. Companion plugs were incubated alone or treated with pyrimidine dimer-specific endonuclease from *Micrococcus luteus*¹⁶ to induce single-strand breaks at all dimer sites. After electrophoresis on alkaline agarose gels⁸, DNA was stained with ethidium bromide, and quantitative electronic image recorded¹⁷ and used to calculate the dimer frequency in each sample⁸. Dimer frequencies were determined using three independent gels for each sample, and the dose responses calculated as described⁹. As the analysis of variance showed no significant difference among independent experiments at each wavelength, data from all gels for a given wavelength were pooled and the method of least squares used to determine the slope of a linear dose response curve and its standard deviation. Time–intensity reciprocity was demonstrated at 275 nm to ensure that values at short wavelengths were not artificially depressed by repair during irradiation.

they may be compared directly without normalization at an arbitrary reference wavelength such as 300 nm (ref. 5). The action spectrum for dimer formation in unshielded DNA is similar in shape to the relative action spectrum for DNA damage presented by Setlow², which has been widely used in studies of the biological consequences of ozone depletion (Fig. 1).

Depletion of stratospheric ozone will increase the intensity of mid-ultraviolet radiation (roughly 280 to 320 nm) reaching the surface of the earth (Fig. 2). The biologically effective ultraviolet delivered by a given solar spectrum is the integral over wavelength of the product of the solar irradiance times the action spectrum for the biological system². As wavelengths greater than about 320 nm are not affected greatly by changes in ozone levels, the larger the contribution of such wavelengths to the basal level of biologically effective ultraviolet, the smaller the relative increase resulting from ozone depletion.

For wavelengths below about 310 nm, the action spectrum for damaging unshielded DNA is greater than those for alfalfa or human skin (Fig. 1). Unless adjusted for local absorption², the spectrum for unshielded DNA will yield higher predicted increases in DNA damage resulting from ozone depletion than the spectra for dimer production in the intact organisms. A generalized plant damage action spectrum^{3–5}, extending from 280 to 313 nm, has also been used to predict increases in biologically effective ultraviolet. This spectrum terminates at 313 nm because no data are available for longer wavelengths. The shape of the generalized plant spectrum is similar to that of the alfalfa

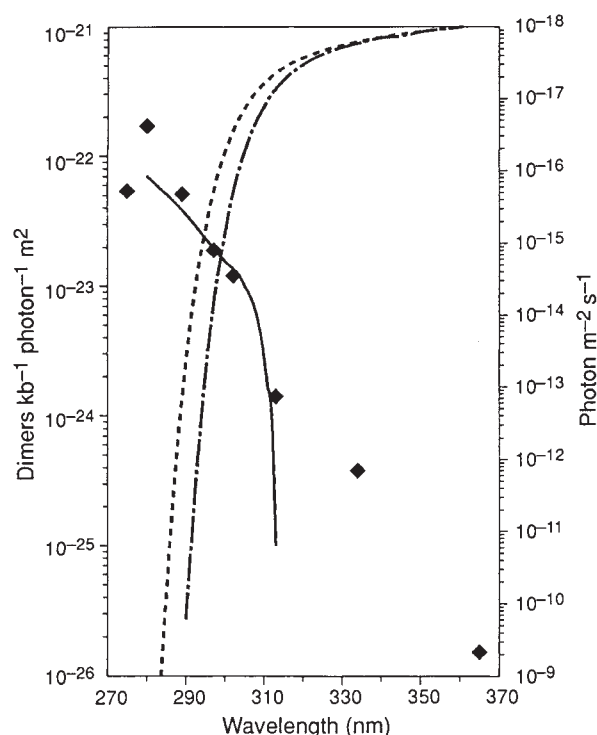


FIG. 2 Comparison of the action spectrum for pyrimidine dimer induction in alfalfa seedlings (♦) with a generalized action spectrum for plant damage^{3–5} (full line). The generalized plant spectrum was normalized to the alfalfa data using a logarithmically weighted least squares procedure at wavelengths from 280 to 302 nm. The value of the generalized plant action spectrum at 313 nm, as originally presented on a linear scale^{4,5}, was 0. Spectral distributions for downward global solar flux at the surface of the earth for a solar angle of 40° and 0.32 cm (broken line) and 0.16 cm (dotted line) of stratospheric ozone¹⁸ show that the most significant changes in ultraviolet due to ozone depletion are for wavelengths less than 320 nm.

spectrum for wavelengths below about 310 nm (Fig. 2). The alfalfa spectrum, however, indicates significant damage by longer wavelengths, which are more intense in solar spectra. Calculations using the generalized plant spectrum do not account for damage produced by wavelengths greater than 313 nm. Such calculations thus underestimate the basal level of DNA damage and, as a result, overestimate the relative increases in biologically effective ultraviolet radiation resulting from ozone depletion, particularly when the intensity at shorter wavelengths is low, that is, at high latitudes.

The action spectrum for dimer induction in alfalfa provides a direct determination of the damage induced by ultraviolet light in DNA of an intact plant. When convoluted with solar spectra, the alfalfa spectrum predicts smaller increases in biologically effective ultraviolet resulting from ozone depletion than calculations based on other widely used action spectra, as shown in Table 2. In contrast to predictions based on the generalized plant action spectrum⁶, the increases in biologically effective ultraviolet radiation determined using the alfalfa action spectrum are little affected by solar angle and, hence, by latitude.

Factors that modify the relative contributions of the shorter and longer wavelength ultraviolet regions will influence the predicted increases in DNA damage resulting from ozone depletion. The alfalfa action spectrum represents the response of seedlings that have minimal levels of prior exposure to ultraviolet and hence of ultraviolet-induced pigments. Such pigments will attenuate shorter ultraviolet wavelengths selectively¹⁴, thus reducing the damage induced by the ultraviolet radiation ($\lambda < 320$ nm) that is elevated by ozone depletion. In contrast, both photosensitized DNA damage and repair of dimers would cause

TABLE 2 Biologically effective ultraviolet enhancement ratios for a 50% O₃ reduction as a function of solar elevation angle and different action spectra

Solar elevation angle	DNA damage action spectrum ²	Generalized plant action spectrum ³⁻⁵	Pyrimidine dimers in human skin ⁸	Pyrimidine dimers in alfalfa
90°	4.49	2.93	2.19	1.74
60°	4.70	3.17	2.28	1.73
40°	5.00	3.75	2.46	1.72
20°	4.77	5.66	2.64	1.68
10°	4.12	7.67	2.56	1.64

Biologically effective ultraviolet enhancement ratios, R_{be} , resulting from reduction of the total ozone column from 0.32 to 0.16 cm are given by the equation

$$R_{be} = \frac{\int_{\lambda} I(\phi, O_3 = 0.16, \lambda) \cdot \sigma(\lambda) \cdot d\lambda}{\int_{\lambda} I(\phi, O_3 = 0.32, \lambda) \cdot \sigma(\lambda) \cdot d\lambda}$$

where I is the downward photon flux at the specified solar elevation angle, ϕ , ozone column density, O_3 , and wavelength, λ , for a given action spectrum, $\sigma(\lambda)$. All integrals were evaluated from 280 to 365 nm. Solar fluxes, at increments of 5 nm, are from ref. 18, with values for wavelengths from 345 to 365 nm determined by extrapolation. Action spectra are from the data in Table 1 or the reference shown in Table 2, with values determined by interpolation at the same 5-nm increments as the solar spectra.

the alfalfa action spectrum to underestimate total DNA damage induced by longer wavelength ultraviolet light ($\lambda > 320$ nm). All three factors would cause our calculations to overemphasize the shorter wavelength ultraviolet radiation and hence overestimate the increase due to lowered ozone levels. Thus, the calculations reported here probably represent a worst-case scenario for the increase in ultraviolet damage of DNA in higher plants resulting from ozone depletion.

Plants are certainly at risk from increased ultraviolet radiation resulting from destruction of stratospheric ozone. In contrast to calculations based on the generalized plant action spectrum⁶, however, our results indicate that plants are not among the most sensitive biological targets. In addition, we find that DNA in plants is no more sensitive to a given level of ozone depletion at middle and high latitudes than at low latitudes.

Note added in proof: Recent reports¹⁹⁻²¹ of the wavelength dependence of the biological effects of ultraviolet light on phytoplankton indicate that the shorter wavelengths account for only a small portion of the total burden and are therefore consistent with our results for alfalfa. □

Received 4 May; accepted 1 July 1992.

- Jagger, J. *Solar-UV Actions on Living Cells* (Praeger, New York, 1985).
- Setlow, R. B. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3363-3366 (1974).
- Caldwell, M. M. *Ecol. Monogr.* **38**, 243-268 (1968).
- Caldwell, M. M. in *Photophysiology* (ed. Giese, A. C.) **6**, 131-177 (Academic, New York, 1971).
- Caldwell, M. M., Camp, L. B., Warner, C. W. & Flint, S. D. in *Stratospheric Ozone Reduction, Solar Ultraviolet Radiation and Plant Life* (eds Worrest, R. C. & Caldwell, M. M.) 87-111 (Springer, Berlin, 1986).
- Madronich, S. *Geophys. Res. Lett.* **19**, 37-40 (1992).
- Sutherland, B. M., Harber, L. C. & Kochevar, I. E. *Cancer Res.* **40**, 3181-3185 (1980).
- Freeman, S. E. *et al. Anal. Biochem.* **158**, 119-129 (1986).
- Quate, F. E., Sutherland, B. M. & Sutherland, J. C. *Appl. theor. Electroph.* **2**, 171-175 (1992).
- Cochill, T. P. in *Topics in Photomedicine* (ed. Smith, K. C.) 1-37 (Plenum, New York).
- Caldwell, M. M., Robberrecht, R. & Flint, S. D. *Physiol. Plant.* **58**, 445-450 (1983).
- Freeman, S. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 5605-5609 (1989).
- Emrick, A. & Sutherland, J. C. *Photochem. Photobiol.* **49**, 355 (1989).
- Tevini, M., Braun, J. & Fieser, G. *Photochem. Photobiol.* **53**, 329-333 (1991).
- Johns, H. E. & Rauth, M. *Photochem. Photobiol.* **4**, 673-692 (1965).
- Carrier, W. L. & Setlow, R. B. *J. Bact.* **102**, 178-186 (1970).
- Sutherland, J. C. *et al. Anal. Biochem.* **163**, 446-457 (1987).
- Shettle, E. P., Nack, M. L. & Green, A. E. S. in *Impacts of Climatic Change on the Biosphere 1: Ultraviolet Radiation Effects* (eds Nachtwey, D. S., Caldwell, M. M. & Biggs, R. H.) 238-249 (National Technical Information Service, Springfield, VA, 1975).
- Helbling, E. W., Villafane, V., Ferrario, M. & Holm-Hansen, O. *Mar. Ecol. Prog. Ser.* **80**, 89-100 (1992).
- Lubin, D. *et al. J. geophys. Res.* **97**, 7817-7828 (1992).
- Ryan, K. G. *J. Photochem. Photobiol.* **13**, 235-240 (1992).

ACKNOWLEDGEMENTS. We thank R. Setlow and T. Cochill for comments, and J. Romeo, D. Monteleone and J. Trunk for technical assistance. Supported by the USDA and USDOE.

Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes

Christian Sturmbauer & Axel Meyer

Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245, USA

SINCE their discovery at the turn of the century¹, the species assemblages of cichlid fishes in the East African Lakes Victoria, Malawi and Tanganyika have fascinated evolutionary biologists. Many models have attempted to account for the 'explosive' evolution of several hundred species within these lakes²⁻⁷. Here we report a case of surprisingly large genetic divergence among populations of the endemic *Tropheus* lineage of Lake Tanganyika. This lineage of six species contains twice as much genetic variation as the entire morphologically highly diverse cichlid assemblage of Lake Malawi and six times more variation than the Lake Victoria species flock. Although it is highly variable in coloration, this group of species and its closest relatives have not undergone appreciable morphological change. The observed geographic pat-

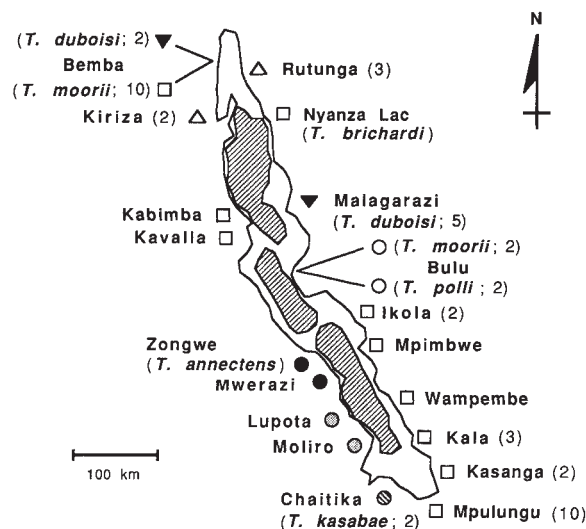


FIG. 1 Localities of 21 *Tropheus* populations from Lake Tanganyika included in this study. Fifty-four individuals, assigned to six currently recognized species by meristic characters (for example, by the number of anal spines) and coloration, were sequenced¹⁰: *Tropheus moorii* Boulenger 1898, *T. annectens* Boulenger 1898, *T. duboisi* Marlier 1959, *T. brichardi* Nelissen and Thys 1976, *T. kasabae* Nelissen 1977 and *T. polli* Axelrod 1978. The populations are identified by the name of the nearest village¹³. The species names are given according to the current taxonomic assignments. Numbers of individuals analysed from each locality are given in parentheses if larger than one. All populations in which the species name was omitted are presently classified as *T. moorii*; this name is only included for those localities in which *T. moorii* is sympatric with other *Tropheus* species. The symbols indicate genetically distinct lineages based on the phylogenetic analyses of mitochondrial DNA sequence data (Fig. 4). The molecular phylogeny is in partial conflict with the present taxonomy. We suggest a revision of the taxonomy of *Tropheus*, which should include molecular data rather than rely solely on taxonomic characters like coloration or the number of spines in the anal fin that are currently used. Stippled areas indicate the locations of the three separate basins during a period of low lake level^{16,17}. In all populations from the geographically widespread lineage for which sequences were determined (squares in Figs 1 and 4) only two substitutions were observed (Fig. 4a) in cytochrome *b* and up to 4% (corrected sequence divergence, see Fig. 3) in the control region. In the central part of the lake two populations from opposite shorelines (Kabimba and Ikola) have a smaller sequence divergence in the control region (0.9%) than to other populations from the same coastline (for example, Kabimba-Bemba=2.1%; Ikola-Nyanza=2.4%).

tern of genetic variation suggests that major lake level fluctuations affected the distribution and speciation of this lineage of cichlid fishes.

Speciation and morphological evolution are two independent processes that usually occur synchronously but may proceed independently⁸. The morphology of 'living fossils', like horseshoe crabs, has remained essentially unchanged for millions of years, although these organisms exhibit normal levels of molecular evolution⁹. In contrast, hundreds of species of cichlid fishes (species flocks) from the African Great Lakes, particularly Lakes Victoria and Malawi, originated extremely rapidly through adaptive radiations and are genetically exceptionally closely related⁷, yet these fishes have undergone major morphological diversifications. The oldest of these lakes, Lake Tanganyika, harbours 171 species assigned to 49 endemic genera which belong to at least seven old, evolutionarily distinct lineages^{10,11}. Because of their long independent evolutionary history, these cichlids are morphologically, ecologically and behaviourally the most diverse species flock of the African lakes^{4,12}.

In most Tanganyikan cichlid lineages, many geographically distinct populations have been described, distinguishable by only minor if any morphological variation but by pronounced differences in coloration¹⁰. The best known examples of this phenomenon are the six described species of the *Tropheus* lineage. Of these, more than 50 distinctly coloured 'races' are currently known, some of which have overlapping distributions¹³. *Tropheus* are strictly confined to rocky habitats for foraging and mating and have limited capacity for dispersal across open water^{13,14}.

Mitochondrial DNA of 54 individuals from 21 populations of all six described species of *Tropheus* was sequenced to investigate the distribution of genetic variation within and among populations and species in relation to taxonomy, coloration and geographic distribution (Fig. 1). Up to 842 base pairs (bp) from the control region and the cytochrome *b* gene were sequenced for each individual (Fig. 2).

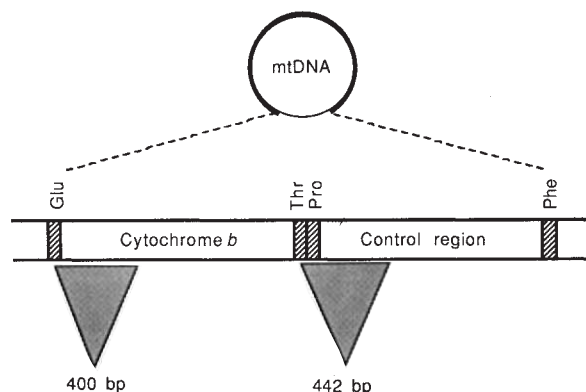


FIG. 2 Portion of the mitochondrial genome showing the sequenced gene segments. A 442-base-pair (bp) segment of the major non-coding region (control region), including a part of the threonine transfer RNA gene (12 bp) and the proline tRNA gene (71 bp), was amplified by the polymerase chain reaction (PCR) and directly sequenced^{7,26} in all 54 individuals. In addition, a 400-bp segment of the cytochrome *b* gene was sequenced for all six species (15 specimens from 13 populations). DNA was extracted from white muscle tissue of frozen or ethanol-preserved specimens. PCR amplifications were done in 25- μ l volumes of TRIS buffer (67 mM, pH 8.8) containing 2 mM $MgCl_2$, 1 mM of each dNTP, 1 μ M of each primer and *Taq* polymerase (1 unit; Cetus). The primers used for the amplification of part of the cytochrome *b* gene were L14724 and H15148, and those for the tRNA and the control region were L15926 and H16498 (details of the protocol and primer sequences have been reported in refs 7, 26). Sequences have been deposited in Genbank under accession numbers Z12047–Z12100 (*Tropheus*, control region); Z12030–Z12044 (*Tropheus*, cytochrome *b*); Z12045 (*Simochromis babaulti*, cytochrome *b*); Z12046 (*Oreochromis tanganyicae*, cytochrome *b*). Swedish Museum of Natural History, NRM 12815. The aligned sequences have also been deposited as supplementary information at the *Nature* office.

The amount of sequence divergence in the mitochondrial control region within the *Tropheus* lineage was compared to that within the cichlid species flocks of Lakes Malawi and Victoria. This comparison revealed a surprisingly high degree of genetic variation in *Tropheus* (Fig. 3). With a maximum corrected sequence divergence¹⁵ (Fig. 3) of 14.5%, the *Tropheus* lineage contains almost twice as much variation as the whole Malawi species flock of about 500 species (8.2%) (ref. 7, and P. Reinthal and A.M. unpublished results) and more than six times as much variation as the Victoria flock of about 300 species (2.3%)⁷. This estimate is also supported by the variation in cytochrome *b*, for which there is up to 5% variation within the *Tropheus* lineage compared with less than 2% in the whole Malawi flock⁷. The species flock of Lake Malawi is estimated to be around 700,000 years old and that of Lake Victoria 200,000

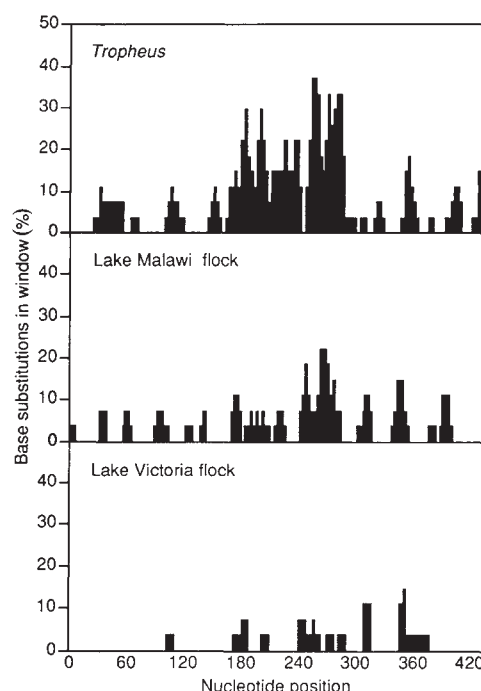


FIG. 3 Sliding window analysis of 442 base pairs (bp) of the mitochondrial control region (window: 9 bp, 3 bp overlap) of *Tropheus* and representative species of the endemic species flocks of lakes Malawi (500–1,000 species) and Victoria (about 300 species)¹². Twenty-one *Tropheus* ($n=54$) populations from Lake Tanganyika were compared with 24 Malawian species from 12 genera ($n=45$) and 14 species of 9 genera ($n=32$) from Lake Victoria⁷. Note the extraordinarily large sequence divergence in *Tropheus*. The genetic variation is given in per cent of the 27 possible (window width, 9 bp) base substitutions from a consensus sequence. Transversion mutations are less frequently observed than transitions among closely related species and therefore trace phylogenetic events more reliably¹⁵. Transversions will cover up previous transitions and estimates of sequence divergence must be corrected for those multiple substitutions. The sequence divergences of the control region were corrected for multiple substitutions according to a formula¹⁵ based on the observed ninefold higher frequency of transitions than transversions. The frequency of transversions was determined by a linear regression, plotting transitions versus transversions of 53 pairwise comparisons of species or populations of up to 8 per cent uncorrected sequence divergence (regression: $y = 0.11 \cdot x + 0.19$, $r^2 = 0.66$, $1/0.11 = 9.1$). The amount of genetic variation within populations was surveyed in two populations. At Bemba, eight out of ten fish were identical; one individual differed by one transition, another by two transitions. At Mpulungu, eight out of ten fish had sequence differences of between 1 and 17 substitutions (3.8%). This population contains several distinct genotypes that are closely related to those found in neighbouring populations at Kasanga, Kala and Wampembe. These findings agree with data from sediment cores from that area, showing the drainage of Mpulungu Bay by two repeated drops in lake level by 400 m 30,000–20,000 years ago²⁷, probably causing admixture of those populations.

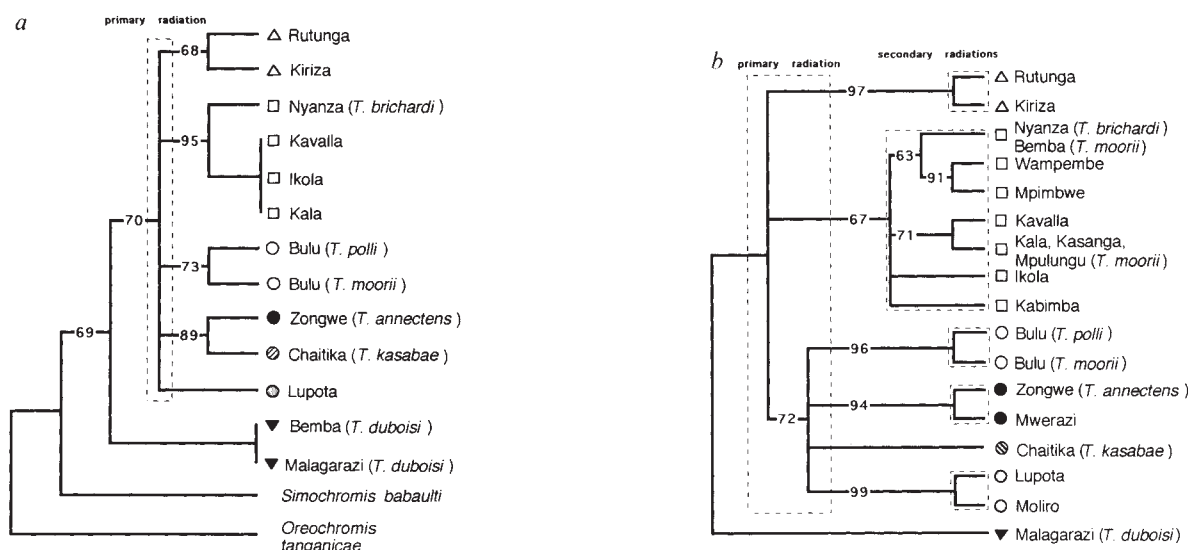


FIG. 4 Molecular phylogeny of the genus *Tropheus*. *a*, Bootstrap consensus tree (100 replications, 50% majority rule) based on cytochrome *b* sequences. *Oreochromis tangerianae*, a member of the tilapia cichlids, a distantly related species to *Tropheus* (and most other endemic Tanganyikan genera (refs 11 and 28; also our unpublished observation), was used as outgroup. Transitions and transversions were equally weighted because of the low frequency of transversions. Thirteen populations of six species from all coastal regions were included in this analysis, one shortest tree was found (heuristic search, ACCTRAN option; tree length, 93; consistency index, 0.75; consistency index excluding uninformative characters, 0.56). *Tropheus duboisi* is the sister group of all other *Tropheus* species; monophyly of the *Tropheus* lineage seems supported by the inclusion of *Simochromis babaulti*, a member of the tribe Tropheini¹⁰. *b*, Bootstrap consensus tree (50% majority rule) based on the control region. Transversions were weighted nine times over transitions, according to their frequency among closely related species (Fig. 3). Three insertions were observed; each was weighted as a single transversion. This analysis included 16 representative populations (heuristic search, 100 replications, ACCTRAN option; weighted tree length, 386; unweighted tree

length, 171; consistency index, 0.61; consistency index excluding uninformative characters, 0.49). To keep the number of taxa manageable for the bootstrap analysis²⁹, each population was represented by only one individual. Parsimony analysis²⁹ of all individuals ($N=36$; individuals with identical sequences were excluded) resulted in 24 equally short trees (heuristic search, ACCTRAN option; weighted tree length, 444; unweighted tree length, 197; consistency index, 0.58; consistency index excluding uninformative sites, 0.5) which agree with the topology shown and varied only slightly in the arrangements of terminal branches. Parsimony analysis and bootstrap procedures were conducted with PAUP (Version 3.0 β ²⁹). The closely related populations from Nyanza and Bemba and the populations from Kala, Kasanga and Mpulungu were each represented by only one individual. The primary radiation and the succeeding more recent secondary radiations of each lineage are indicated by boxes. A neighbour-joining analysis³⁰ (54 individuals) with weighted distance matrix (weighing transversions and gaps nine times over transitions) resulted in a tree in agreement with the branching order shown here.

years⁷. The geological ages of the lakes themselves are gauged to be somewhat greater^{16,17}. On the basis of these estimates and the assumption of comparable rates of molecular divergence among the members of each of the African species flocks, the *Tropheus* lineage might therefore be about 1.25 million years old. Although the absolute ages of these species assemblages remain approximations, the relative divergence demonstrates the antiquity of the *Tropheus* lineage compared to other morphologically highly diverse species assemblages.

Tropheus is morphologically distinct from the other genera (for example, *Simochromis babaulti*; Fig. 4a) assigned to the same tribe, Tropheini¹⁰. Despite the large amount of sequence divergence and inferred old age, all species of *Tropheus* are morphologically largely identical even where they live sympatrically. Genetically, *Tropheus duboisi* is the most divergent species of the genus, but morphologically it differs only slightly from all other *Tropheus* by a narrower mouth and slight dentitional differences¹⁸.

The extreme levels of morphological diversity among the main lineages of the Tanganyikan species flock, which lack intermediate morphotypes, led Greenwood¹² to hypothesize that these lineages were formed rapidly after the colonization of the lake by several ancestors filling available niches, but then remained morphologically unchanged. Lacking crucial fossil data, this scenario had remained untested but seems now to be supported by the molecular data for *Tropheus*. Among various possible explanations for stasis in morphology^{8,19–21}, the action of stabilizing selection, as a consequence of species as part of 'tightly packed' communities²¹, might best explain the situation

in the *Tropheus* lineage. Future molecular investigation of other Tanganyikan lineages might determine whether the morphological stasis and the mode of evolution inferred from the *Tropheus* lineage is a general attribute of the species flock of Lake Tanganyika.

The evolutionary relationships among *Tropheus* were reconstructed based on DNA sequences. The phylogenetic information obtained from the more slowly evolving cytochrome *b* gene indicated the relationship of *Tropheus* to morphologically dissimilar genera (such as *Simochromis*; Fig. 4a). The rapidly evolving control region resolved more recent phylogenetic events within the genus (Fig. 4b). *Tropheus duboisi* appears to be the closest living relative of all other *Tropheus* species. After this first split of the *Tropheus* line, two rapid creations of lineages (radiations) occurred. In a primary radiation (Fig. 4a, b) probably all rocky shores of the lake were colonized. Six lineages with similar genetic divergence (Fig. 4b) originated during the primary radiation; they are separated by an average corrected sequence divergence in the control region of 8.2% (standard deviation 1.4%; 314 pairwise comparisons). Each of these lineages has undergone a 'secondary radiation' (Fig. 4b).

These secondary radiations might have been triggered by a large fluctuation in lake level. The similar sequence divergences among members of each of these radiations (average in the control region: 2.7%, standard deviation 1.0%, 195 pairwise comparisons between populations from each lineage) of all the secondary radiations suggests that they might have arisen at a similar time in the geological history of the lake. Seismic data show that about 200,000 (ref. 16) to 75,000 (ref. 17) (C. A.

Scholz, personal communication) years ago the level of Lake Tanganyika dropped 600 m below its present level, probably splitting the lake into three sublakes for several tens of thousands of years (Fig. 1). This event must have had trenchant effects on the biogeographic distribution of many species and probably led to widespread extinctions and fusions of formerly isolated populations. Because *Tropheus* are confined to rock habitats^{14,18} many populations might have disappeared in shallow areas of the lake. Populations at rocky steeply sloping shores probably survived the drop in lake level intact.

The geographic distribution of five of the observed six primary lineages is limited, but members of one lineage (square symbols in Figs 1 and 4) are widespread throughout the lake. They have colonized newly available habitats, some of which are geographically separate and were probably previously occupied by other *Tropheus* lineages (the southeastern part of the lake for example). The distribution of genetically closely related populations on both sides of the lake (Fig. 1) suggests that colonizations not only took place along the same side of the lake, but these fish seemed to have crossed from one shore to the other. The northern and southern shorelines of the sublakes during the low lake level might have permitted the crossing of the lake. The molecular divergence data suggest that abiotic factors, such as major lake level fluctuations, might be partly responsible for speciation in *Tropheus*.

The DNA data show that genetically closely related populations can differ considerably in coloration. For example, the two sympatric, genetically closely related species (3.1% sequence divergence) at Bulu are remarkably differently coloured. Behavioural observations demonstrated the importance of coloration during conspecific interactions^{14,22}. Sexually and socially selected traits might undergo more rapid divergence than might traits that are under natural selection^{3,6,23,24}. Owen *et al.*²⁵ reported on extremely young (estimated to be less than 1,000 years old), colorationally distinct species from the Lake Malawi cichlid species flock. *Tropheus* may be a model of how rapidly differences in sexually selected traits such as coloration might arise as reproductive barriers (without concordant morphological diversification) during periods of isolation, possibly aiding the 'explosive' pace of speciation in East African cichlid species flocks. □

Received 20 February; accepted 16 June 1992.

1. Boulenger, C. L. *Trans. Zool. Soc. Lond.* **15**, 1–30 (1898).
2. Avise, J. C. *Nature* **347**, 512–513 (1990).
3. Kosswig, C. *Nature* **159**, 604–605 (1947).
4. Fryer, G. & Iles, T. D. *The Cichlid Fishes of the Great Lakes of Africa* (Oliver & Boyd, Edinburgh, 1972).
5. Mayr, E. in *Evolution of Fish Species Flocks* (eds Echelle, A. A. & Kornfield, I.) 3–11 (University of Maine Press, Orono, 1984).
6. Dominey, W. J. in *Evolution of Fish Species Flocks* (eds Echelle, A. A. & Kornfield, I.) 231–249 (University of Maine Press, Orono, 1984).
7. Meyer, A., Kocher, T. D., Basasibwaki, P. & Wilson, A. C. *Nature* **347**, 550–553 (1990).
8. Larson, A. in *Speciation and its Consequences* (eds Otte, D. & Endler, J. A.) 579–598 (Sinauer, Sunderland, Massachusetts, 1989).
9. Selander, R. K., Yang, S. Y., Lewontin, R. C. & Johnson, W. E. *Evolution* **24**, 402–414 (1970).
10. Poll, M. *Ac. Roy. de Belg. Memoires de la classe des sciences. Collection in 8°-2e serie, T.XLV, Fasc. 2*, 1–16 (1986).
11. Nishida, M. *Experientia* **47**, 974–979 (1991).
12. Greenwood, P. H. in *Evolution of Fish Species Flocks* (eds Echelle, A. A. & Kornfield, I.) 141–154 (University of Maine Press, Orono, 1984).
13. Brichard, P. *Cichlids of Lake Tanganyika* (TFH Publications Inc., Neptune City, New Jersey, 1989).
14. Sturmbauer, C. thesis, Univ. Innsbruck, Austria (1990).
15. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. *J. molec. Evol.* **18**, 225–239 (1982).
16. Tiercelin, J.-J., Mondeguer, A. in *Lake Tanganyika and its Life* (ed. Coulter, G. W.) 7–48 (Oxford Univ. Press, London, 1991).
17. Scholz, C. A. & Rosendahl, B. R. *Am. Assoc. Petrol. Geol. Memoir* **50**, 151–168 (1990).
18. Marlier, G. *Rev. Zool. Bot. Afr.* **59**, 164–183 (1959).
19. Charlesworth, B., Lande, R. & Slatkin, M. *Evolution* **36**, 474–498 (1982).
20. Charlesworth, B. & Lande, R. *Nature* **296**, 610–611 (1982).
21. Turner, J. R. in *Patterns and Processes in the History of Life* (eds Raup, D. M. & Jablonski, D.) 183–207 (Springer, Heidelberg, 1986).
22. Neilsen, M. *Rev. Zool. Bot. Afr.* **90**, 17–29 (1976).
23. West-Eberhard, M. J. *Q. Rev. Biol.* **58**, 155–183 (1983).
24. Coyne, J. A. *Nature* **355**, 511–515 (1992).
25. Owen, R. B. *et al. Proc. R. Soc. B* **240**, 519–553 (1990).
26. Kocher, T. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6196–6200 (1989).
27. Gasse, F., Ledee, V., Massault, M. & Fontes, J. C. *Nature* **342**, 57–59 (1989).
28. Kornfield, I. in *Cichlid Fishes: Behavior, Ecology and Evolution* (ed. Keeley, M. H. A.) 103–128 (Chapman & Hall, London, 1991).
29. Swofford, D. L. *Phylogenetic Analysis Using Parsimony, Version 3.0r* (Illinois Natural History Survey, Champaign, 1991).

30. Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406–425 (1987).

SUPPLEMENTARY INFORMATION. Requests should be addressed to the London editorial office of *Nature*. ACKNOWLEDGEMENTS. We thank L. DeMasson and S. Kullander for some of the samples, the Government of the Republic of Burundi, especially M. Boniface Nyakagani, for research permission, and D. Dykhuizen, D. Futuyma, J. Levinton, E. Mayr, G. Orti, T. Titus, P. Wilson and K. Wollner for discussion. A.M. is supported by the NSF and C.S. by the Austrian Science Foundation, through an Erwin Schroedinger Postdoctoral Fellowship.

Defective regulation of outwardly rectifying Cl⁻ channels by protein kinase A corrected by insertion of CFTR

Marie Egan*, Terence Flotte*†, Sandra Afione†, Rikki Solow†, Pamela L. Zeitlin*, Barrie J. Carter† & William B. Guggino*‡

* Departments of Physiology and Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

† Laboratory of Molecular and Cellular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, USA

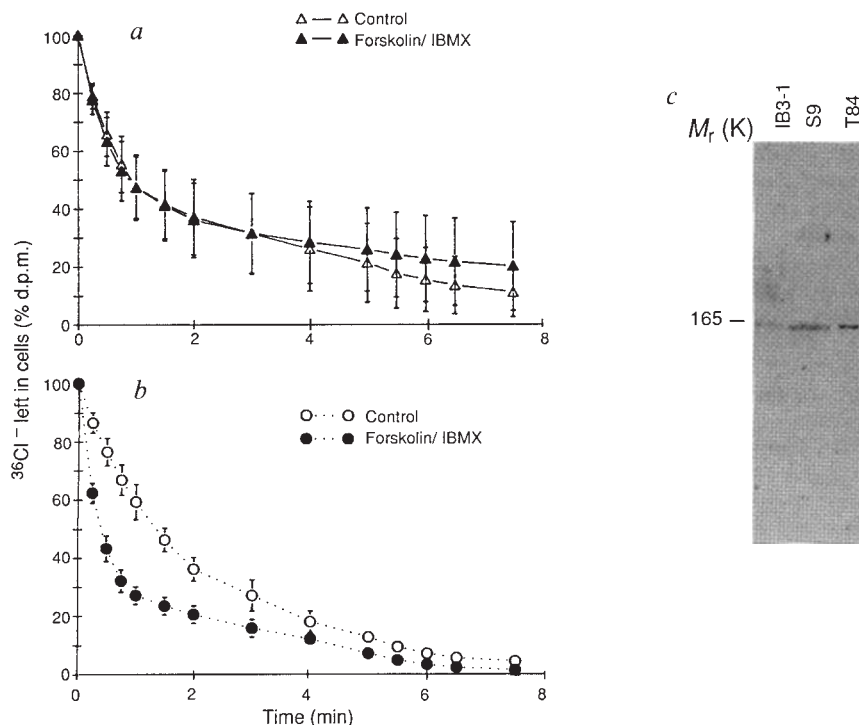
CYSTIC fibrosis (CF) is a lethal genetic disease resulting in a reduced Cl⁻ permeability¹, increased mucous sulphation², increased Na⁺ absorption³ and defective acidification of lysosomal vesicles⁴. The CF gene encodes a protein (the cystic fibrosis transmembrane conductance regulator, CFTR⁵) that can function as a low-conductance Cl⁻ channel with a linear current-voltage relationship whose regulation is defective in CF patients^{6–8}. Larger conductance, outwardly rectifying Cl⁻ channels are also defective in CF and fail to activate when exposed either to cyclic AMP-dependent protein kinase A or to protein kinase C^{9–13}. The role of the outwardly rectifying Cl⁻ channel in CF has been questioned¹⁴. We report here that expression of recombinant CF genes using adeno-associated virus vectors in CF bronchial epithelial cells corrects defective Cl⁻ secretion, that it induces the appearance of small, linear conductance Cl⁻ channels, and restores protein kinase A activation of outwardly rectifying Cl⁻ channels. These results re-establish an involvement of outwardly rectifying Cl⁻ channels in CF and suggest that CFTR regulates more than one conductance pathway in airway tissues.

Excised patch clamp experiments were done on a CF bronchial epithelial cell line (IB3-1) obtained from a CF patient with severe CF¹⁵ or on IB3-1 cells either cotransfected¹⁶ with adeno-associated virus/cystic fibrosis transmembrane regulator (AAV-CFTR) and AAV-*neo* vector plasmids (S9, and C38 cells) or infected with AAV-CFTR viral particles (A0, and 2F2 cells). AAV-CFTR plasmids (pSA315 or pSA306) were cotransfected into IB3-1 with the AAV-*neo* plasmid to produce geneticin-resistant clones (S9 or C38, respectively) and screened in a ³⁶Cl⁻ efflux assay (Fig. 1a, b). Stimulation of ³⁶Cl⁻ efflux by forskolin and 3-isobutyl-1-methylxanthine (IBMX) did not occur in the uncorrected IB3-1 cells (Fig. 1a), but was evident in cells containing recombinant AAV-CFTR vectors (Fig. 1b). These results verify that expression of normal CFTR in CF cells (Fig. 1c) confers forskolin-dependent stimulation of Cl⁻ secretion^{17,18}. Transfection experiments with a plasmid (pSA464) with a frame-shift mutation at nucleotide 993, with subsequent termination at 1,111, gave no complemented clones. When the AAV-CFTR vector pSA306 was packaged into AAV particles and infected into IB3-1 cells, cAMP regulation of Cl⁻ efflux was complemented in the mass culture (A0 cells). However, when IB3-1 cells were transduced with AAV particles containing the frame-

‡ To whom correspondence should be addressed.

FIG. 1 Complementation of CF airway cells. Representative Cl^- efflux measurements in IB3-1 (a) and S9 (b) cells. Data points represent mean $\pm$ s.e.; $n=3$ for all points except IB3-1 treated with IBMX (100 μM) and forskolin (10 μM); $n=5$. In addition, forskolin-dependent increases in $^{36}\text{Cl}^-$ efflux were also noted in C38 ($n=3$) and A0 ($n=10$), but not in 2F2 ($n=4$). c, Protein levels in CF airway cells. Western blot using anti-CFTR, pre-ATP binding domain peptide antisera²³. Whole cell extract (180 μg) in lanes 1–2, and 90 μg in lane 3. The 165K band has been shown previously to be CFTR. Note increased CFTR protein levels in S9 compared to IB3-1 at a similar molecular mass as in T84 cells.

METHODS. Cells were grown in tissue culture¹³ and $^{36}\text{Cl}^-$ assay done at 37 °C as described previously²⁴. In pSA315 (in clone S9) the full-length CFTR cDNA¹⁸ is inserted downstream of a 60-base-pair synthetic oligonucleotide contained between the AAV ITR (inverted terminal repeat) regions, and pSAS306 (C38 or A0 cells) is analogous except for a deletion of the first 350 nucleotides of the CFTR cDNA. pSA464 (in 2F2 cells) contains a frameshift mutation at nucleotide 993 which prevents functional CFTR expression (T.F. *et al.*, manuscript, in preparation). Expression of CFTR from the vector in A0, C38, and S9 cells was confirmed both by immunofluorescence and immunoblotting techniques. The presence of AAV-CFTR was determined in A0, 2F2, S9 and C38 cells by Southern blotting and expression of messenger RNA from AAV-CFTR by primer extension in C38 cells (ref. 25, and T.F. *et al.*, manuscript in preparation). For western blotting, protein was electrophoresed on a 5% SDS-PAGE system. The gel was transferred to nitrocellulose and



incubated with pre-nucleotide-binding fold antiserum as previously described²³. Bands were detected using enhanced chemiluminescence.

shift mutant vector SA464 (2F2 cells; see Fig. 1 legend), similar complementation did not occur.

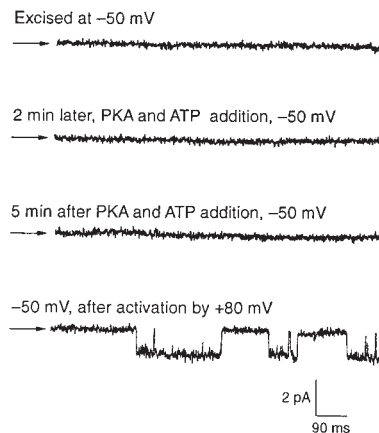
No spontaneous channel activity occurred when patches were excised from IB3-1 cells at -50 mV and at 22 °C ($n=0/10$: number of activations/number verified to have outwardly rectifying Cl^- channels by voltage activation). Addition of protein kinase A (PKA) and ATP two minutes after excision from IB3-1 cells failed to activate channels for as long as 2–5 min. Subsequent large depolarizations between $+70$ and $+100$ mV verified that outwardly rectifying Cl^- channels were present ($n=0/6$, see Fig. 2). PKA also failed to activate Cl^- channels in IB3-1 cells infected with AAV virus particles containing CFTR complementary DNA with a frameshift mutation that cannot produce a functional CFTR (2F2 cells), but strong depolarizations did confirm the presence of outwardly rectifying Cl^- channels ($n=0/3$). Thus, PKA does not activate Cl^- channels in cells from CF patients, consistent with previous data^{7–11}.

Complemented IB3-1 cells (S9, A0 and C38) were pretreated with 100 μM IBMX and 10 μM forskolin for 20 min at 22 °C. Two types of channels were active in pretreated cells immedi-

ately on patch excision at -50 mV (Fig. 3a; $n=36$ including: S9, $n=12$; A0, $n=14$; and C38, $n=10$). In 23 patches, only the small linear conductance ion channels were active. Subsequent depolarization of these 23 patches verified that outwardly rectifying Cl^- channels were not present. But in 13 patches both linear and outwardly rectifying Cl^- channels were active on excision at -50 mV (without subsequent depolarization), suggesting that outwardly rectifying Cl^- channels were activated by pretreatment with IBMX and forskolin (Fig. 3a). Small linear Cl^- channel activity ran down rapidly (within 1–60 s) after excision, as shown previously¹⁹, whereas outwardly rectifying channels did not. Low-conductance channel activity could be restored by adding ATP in the bath solution ($n=6$ including: S9, $n=3$; C38, $n=3$, Fig. 3b). Representative current versus voltage relationships for both channels are shown in Fig. 3c. Outwardly rectifying chloride channels were identical to those reported previously¹³. Outwardly rectifying Cl^- channels were sensitive to as little as 50 μM (4,4'-diisothiocyanostilbene-2,2'-disulphonic acid (DIDS)) in bath solution ($n=3$), whereas low-conductance channels were insensitive to DIDS at con-

FIG. 2 Chloride channels in CF bronchial epithelial cells (CFBE). Excised patches from IB3-1 parent cells were held at -50 mV for 2 min before addition of 50 nM PKA and 1 mM Mg-ATP to the cytoplasmic side of the membrane. The failure of PKA activation in this patch was monitored for an additional 5 min. Presence of outwardly rectifying chloride channels was verified by voltage activation.

METHODS. Data were recorded by an EPC-7 electrometer (List Darmstadt, Germany), filtered at 1 kHz (Frequency Devices, Haverhill, Massachusetts), digitized with a pulse code modulator (Sony Corporation), stored on a video cassette recorder (Japanese Victor Corporation), and were analysed with an Axolab 1100 interface, an AST Premium/386 computer (AST Research Inc., Irvine, California) and pClamp software (Axon Instruments, Foster City, California). The bath contained 150 mM NaCl, 2 mM MgCl_2 , 1 mM EGTA, 5 mM HEPES and 0.5 mM CaCl_2 (free calcium was roughly 100 nM). Pipette solution was 150 mM NaCl, 2 mM CaCl_2 , 5 mM HEPES. The pH of all solutions was 7.3. Arrows indicate the closed state. For all figures the data were filtered at 300 Hz.



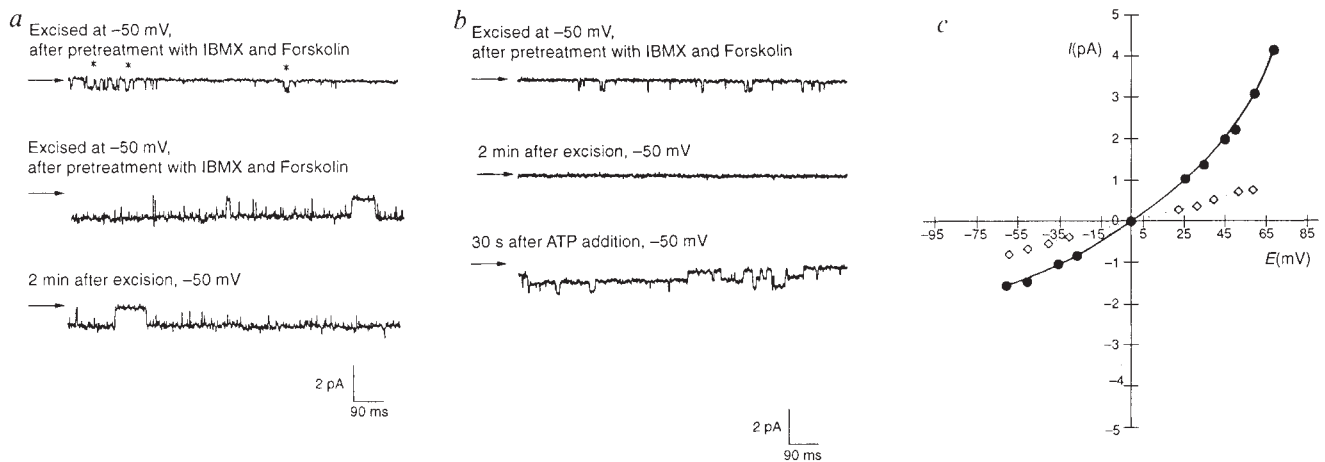


FIG. 3 Activation of Cl^- channels in complemented IB3-1 cells. Cells were pretreated with IBMX and forskolin before excision. *a*, Excised patches were held at -50 mV. Both low-conductance channels (asterisks indicate low-conductance channel activity) and outwardly rectifying Cl^- channels were present in the patch. *b*, Effects of ATP on low-conductance Cl^- channels. Rundown of low-conductance channel activity occurred within 850 ms of patch excision. Two minutes after excision, 1 mM ATP was added to the bath and reactivated low-conductance Cl^- channels. Only low-conductance channels are shown. *c*, Current versus voltage relationships of low conduct-

ance and outwardly rectifying channels. ●, Outwardly rectifying and ◇, low-conductance channels. Pipette and bath solutions as in Fig. 2. Anion permeability sequence for the low-conductance channel was $\text{Br}^- > \text{Cl}^- > \text{I}^-$. For the low-conductance channels, the reversal potential was $+15$ mV in asymmetric solutions (150 mM Cl^- pipette per 300 mM Cl^- bath), consistent with a Cl^- current (expected $= +16.7$ mV). The average single-channel conductance was 13 ± 0.5 pS ($n=3$). Experiments depicted in *a-b* were done on C38 cells and in Fig. 3c on S9 cells.

centrations as high as $500 \mu\text{M}$ ($n=3$). DIDS sensitivity is characteristic of outwardly rectifying Cl^- channels²⁰, in contrast to the low-conductance channels, which are relatively insensitive to DIDS²¹.

No spontaneous channel activity was noted for at least 5 min in patches excised at 22°C and at -50 mV from A0, S9 and C38 cells not pretreated with agonists ($n=0/23$). After an initial 2–5 min period to verify that spontaneous activation had not occurred, the addition of 50 nM PKA and 1 mM ATP activated both small linear conductance ion channels and outwardly rectifying Cl^- channels (Fig. 4; $n=17$, including: S9, $n=8$; C38, $n=6$; A0, $n=3$) within 2–5 min. The activation of outwardly rectifying Cl^- channels by PKA and ATP was identical to our previous studies in normal airway cells¹³. This demonstrates that complementation with AAV-CFTR vectors corrects defective regulation of outwardly rectifying Cl^- channels and induces the expression of small conductance, linear channels.

It has been suggested that activation of outwardly rectifying Cl^- channels could be caused by inadvertent activation¹⁴. This is extremely unlikely for the following reasons. First, when patches are excised at 22°C from unstimulated normal airway cells, neither PKA nor ATP by themselves induce activation. Also, PKA and ATP are not effective in the presence of the Walsh PKA inhibitor¹³. Likewise, PKC activation does not occur with individual cofactors alone (ATP and dioctanoylglycerol) and can be inhibited by staurosporine¹³. These experiments provide convincing evidence that a phosphorylation-dependent reaction plays a role in the activation mechanism. Second, in patches excised at 22°C and at -50 mV either from unstimulated normal cells (ref. 13, and this study) or from CF cells either unstimulated or treated with IBMX and forskolin, or from IB3-1 cells treated with AAV-CFTR containing a frameshift mutation that prevents functional CFTR expression (in this study 40 patches were verified to possess outwardly rectifying Cl^- channel and none displayed spontaneous activation at 22°C and -50 mV), spontaneous activation does not occur. Clearly, these experiments eliminate the possibility of mistaking inadvertent activation for PKA activation.

Outwardly rectifying Cl^- channels can also be activated by depolarizing voltages, trypsin, and higher temperatures (see refs 13, 22). We have observed that excising patches from unstimu-

lated cells at 37°C does cause spontaneous activation in about one quarter of the patches ($n=2$ out of 8). In contrast, at 22°C activation only occurs in prestimulated cells expressing recombinant CFTR. Welsh and collaborators²² have studied the role of different activators and concluded that many aspects of outwardly rectifying Cl^- channel activation are normal in CF, that only phosphorylation-dependent activation is defective, and that temperature, voltage and proteolysis do not have any physiological role in activating these channels.

How does CFTR control PKA activation of outwardly rectifying Cl^- channels? One intriguing possibility is that the CF gene also encodes this channel. We think that this is unlikely for several reasons. First, CF gene expression does not seem to

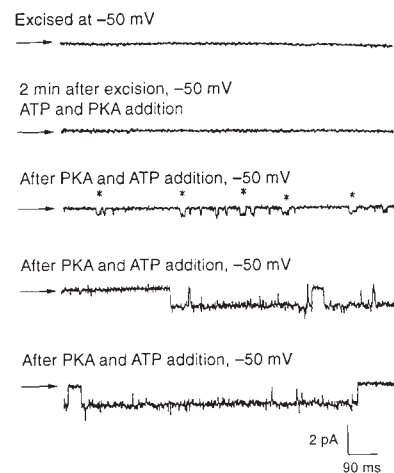


FIG. 4 PKA-activation of low-conductance and outwardly rectifying Cl^- channels. Patches excised from S9 cells not pre-stimulated with agonists and held at -50 mV for at least 2 min before PKA and Mg-ATP were added (see Fig. 2). Asterisks indicate low-conductance channel activity. Bath and pipette solutions as in Fig. 2. In most experiments involving PKA activation, low-conductance channels activated before outwardly rectifying Cl^- channels, although in some instance both channels could be observed to activate simultaneously.

correlate with the presence of outwardly rectifying Cl^- channels¹⁴. Second, CFTR introduced into experimental cell systems such as COS cells, SF9 cells and *Xenopus* oocytes⁶⁻⁸ expresses a small, linear conductance Cl^- with blocker sensitivities and ionic selectivities very different from the larger channel. Another compelling possibility is that CFTR can have more than one function⁶. For example, CF is associated with defective acidification of intracellular organelles⁴. Thus, it is possible that CFTR may induce PKA regulation by affecting outwardly rectifying Cl^- channels as they are processed in the intracellular organelles. Alternatively, CFTR may associate with or transport a substance that is necessary for PKA activation of outwardly rectifying Cl^- channels⁴. □

Received 29 April; accepted 25 June 1992.

1. Frizzell, R. A. *Trends Neurosci.* **10**, 190-193 (1987).
2. Cheng, P. W. et al. *J. clin. Invest.* **84**, 68-72 (1989).
3. Quinton, P. M. *FASEB J.* **4**, 2709-2717 (1990).
4. Barasch, J. et al. *Nature* **352**, 70-73 (1991).
5. Riordan, J. R. et al. *Science* **245**, 1066-1073 (1989).
6. Kartner, N. et al. *Cell* **64**, 681-691 (1991).
7. Anderson, M. P. et al. *Science* **253**, 202-205 (1991).
8. Bear, C. E. et al. *J. biol. Chem.* **266**, 19142-19145 (1991).
9. Li, M. et al. *Nature* **331**, 358-360 (1988).
10. Li, M. et al. *Science* **244**, 1353-1356 (1989).
11. Schoumacher, R. A. et al. *Nature* **330**, 752-754 (1987).
12. Chen, J. H., Schulman, H. & Gardner, P. *Science* **243**, 657-660 (1989).
13. Hwang, T. C. et al. *Science* **244**, 1351-1353 (1989).
14. Ward, C. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5277-5281 (1991).
15. Zeitlin, P. L. et al. *Am. J. respir. Cell molec. Biol.* **4**, 313-319 (1991).
16. Flotte, T. R. et al. *Am. J. respir. Cell molec. Biol.* (in the press).
17. Rich, D. P. et al. *Nature* **347**, 775-784 (1991).
18. Drumm, M. L. et al. *Cell* **62**, 1227-1233 (1990).
19. Anderson, M. P. et al. *Cell* **67**, 775-784 (1991).
20. Bridges, R. J., Worrell, R. T., Frizzell, R. A. & Benos, D. J. *Am. J. Physiol.* **256**, C902-C912 (1989).
21. Anderson, M. P. & Welsh, M. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6003-6007 (1991).
22. Welsh, M. J., Li, M. & McCann, J. D. *J. clin. Invest.* **84**, 2002-2007 (1989).
23. Zeitlin, P. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 344-347 (1992).
24. Trapnell, B. C. et al. *J. biol. Chem.* **266**, 10319-10323 (1991).
25. Flotte, T. et al. *Ped. Pulm.* **6** (suppl.) 247 (1991).

ACKNOWLEDGEMENTS. Full-length CFTR cDNA was a gift from M. Drumm. Research supported by the NIH and the Cystic Fibrosis Foundation.

High brain densities of the immunophilin FKBP colocalized with calcineurin

Joseph P. Steiner, Ted M. Dawson*, Majid Fotuhi, Charles E. Glatt, Adele M. Snowman, Noam Cohen & Solomon H. Snyder†

Departments of Neuroscience, Pharmacology and Molecular Sciences, Psychiatry and Behavioral Sciences, and *Neurology, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

THE immunophilins cyclophilin and FK506 binding protein (FKBP) are small, predominantly soluble proteins that bind the immunosuppressant drugs cyclosporin A and FK506, respectively, with high affinity, and which seem to mediate their pharmacological actions^{1,2}. The Ca^{2+} -dependent protein phosphatase, calcineurin, binds the cyclophilin-cyclosporin A and FKBP-FK506 complexes, indicating that calcineurin might mediate the actions of these drugs³. A physiological role for the immunophilins in the nervous system is implied by a close homology between the structure of NINA A, a protein in the neural retina of *Drosophila*, and cyclophilin^{4,5}, as well as by the high density of FKBP messenger RNA in brain tissue⁶. Here we report that the levels of FKBP and mRNA in rat brain are extraordinarily high and that their regional localization is virtually identical to that of calcineurin, indicating that

there may be a physiological link between calcineurin and the immunophilins. We also show that at low concentrations FK506 and cyclosporin A enhance the phosphorylation of endogenous protein substrates in brain tissue and in intact PC12 cells, indicating that these drugs may inhibit phosphatase activity by interacting with the immunophilin-calcineurin complexes.

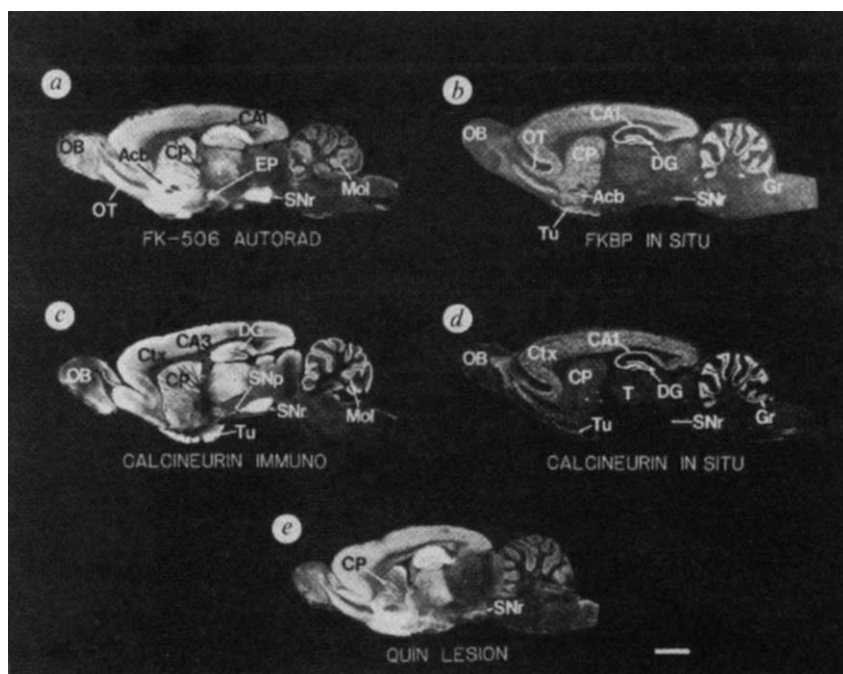
[³H]FK506 binds saturably and with high affinity to soluble and particulate fractions of rat brain (data not shown). In peripheral tissues in both membrane and soluble brain fractions, a high-affinity binding site has a K_D of about 0.6 nM. We also observe a lower-affinity site in both soluble and particulate fractions with a K_D of 30 nM. The immunosuppressant rapamycin has a K_i value of about 1 nM, whereas cyclosporin A fails to inhibit binding at concentrations as high as 100 μM . The density of [³H]FK506 binding sites in various brain regions greatly exceeds levels in peripheral tissues (Table 1). Regions such as the striatum and hippocampus have binding levels about 15 times greater than the highest levels of any of the peripheral tissues, including organs involved in immune responses such as thymus and spleen. Marked regional differences are evident in the brain. Soluble binding in the cerebellum is less than one-tenth the levels in the striatum. Binding is also low in the brain stem and hypothalamus. Although FKBP has been reported as being primarily a soluble protein, we find comparable levels of [³H]FK506 binding in soluble and membrane fractions in the brain. By contrast, in peripheral tissues soluble levels of binding are always higher than membrane values, although in the thymus and spleen particulate binding is about half the soluble levels. Scatchard analysis and drug specificity studies indicate that the properties of [³H]FK506 binding sites are closely similar in membrane and soluble fractions in all brain regions and peripheral tissues (data not shown). Interestingly, in soluble fractions from all brain regions and in membrane fractions from most brain areas, only a single high-affinity binding site with a K_D of about 0.6 nM is detected. But the cerebral cortex, corpus striatum, hippocampus, thymus and spleen membrane fractions give biphasic Scatchard plots with the second low affinity site having a K_D of about 30 nM.

Autoradiograms of [³H]FK506 binding sites and *in situ* hybridization of mRNA for FKBP highlight striking regional variations (Fig. 1). The regional patterns for *in situ* hybridization and autoradiography are essentially the same, ensuring that [³H]FK506 labels FKBP and that the FKBP distribution is essentially the same as that of its mRNA. Highest numbers of [³H]FK506 binding sites are evident in the corpus striatum and the associated nucleus accumbens, and comparable levels are evident in the substantia nigra zona reticulata. The entopeduncular nucleus and the striatonigral pathway are labelled, suggesting that labelling in the nigra reflects the descending striatonigral pathway. We have confirmed this by quinolinate lesions of the corpus striatum, which reduce binding in the striatum and abolish it in the substantia nigra (Fig. 1). High [³H]FK506 binding and FKBP mRNA are also apparent in the hippocampus, with particular enrichment in CA1, and lower amounts in CA3 and the dentate gyrus. [³H]FK506 binding in the cerebral cortex is most prominent in superficial layers. Very high densities of [³H]FK506 binding are also evident in the olfactory tubercle and olfactory tract. In the cerebellum, FKBP mRNA is most enriched in the granule cell layer and [³H]FK506 binding sites are most abundant in the molecular layer, indicating that FKBP may occur in parallel fibres arising from granule cells.

Calcineurin immunoreactivity and mRNA are very similar in their localizations to FKBP. High densities are evident in the corpus striatum, striatonigral pathway and the substantia nigra zona reticulata. In the hippocampus, high densities are most apparent in the CA1 region. Within the cerebral cortex, calcineurin is also enriched in superficial layers. The olfactory tubercle as well as the olfactory tract display very high levels of calcineurin. As with FKBP, cerebellar calcineurin mRNA is most concentrated in the granule cell layer and the protein is

† To whom correspondence should be addressed.

FIG. 1 Dark-field photomicrographs illustrating the dramatic colocalization of FKBP, as labelled by [3 H]FK506 autoradiography (a) with calcineurin immunoreactivity (c). Serial sagittal sections of rat brain, 20- μ m thick, were labelled with [3 H]FK506 by preincubating (60 mins) slide-mounted tissue sections in a buffer consisting of 50 mM HEPES, 2 mg ml $^{-1}$ BSA, 0.02% Tween-20, pH 7.4. This was followed by incubating (60 mins) the tissue sections in 50 mM HEPES, 2 mg ml $^{-1}$ BSA, 0.02% Tween-20, pH 7.4, buffer containing 1 nM [3 H]FK506 (86.5 Ci mmol $^{-1}$; DuPont-NEN, Boston). Nonspecific binding was defined by the addition of 1 μ M FK506. Tissue was then rinsed for 4 $\times$ 5 min in fresh, ice-cold buffer before drying. Radiolabelled sections were then juxtaposed to tritium-sensitive film. Under our conditions, nonspecific binding was routinely at background levels. Immunohistochemistry was as follows: free-floating 40- μ m thick tissue sections from adult male Sprague-Dawley rats perfused with 4% freshly depolymerized paraformaldehyde in 0.1 M phosphate buffer were incubated overnight in Tris-buffered saline (50 mM Tris-HCl) containing affinity purified calcineurin antiserum (1:50) (gift from C. Klee). Sections were stained with an avidin-biotin-peroxidase system (Vector Labs) with diaminobenzidine as the chromogen. In addition, the mRNA for both FKBP (b) and calcineurin (d) have the same distribution. *In situ* hybridization was as described²⁵ using serial sections to those in a. Antisense oligonucleotides were end-labelled with [35 S]dATP. For FKBP, the following antisense oligonucleotides, complementary to nucleotides 70–114, 214–258 and 441–485 of the cloned cDNA^{6,26} were used. For calcineurin, the following antisense oligonucleotides, complementary to the nucleotides of the cloned cDNA (1,363–1,410, 1,711–1,758) (ref. 27) and 1,339–1,386, 1,569–1,616 (ref. 28) were used. Quinolinic acid lesions within the striatum (e) reduce [3 H]FK506 binding, demonstrating that FKBP is localized to the striatonigral pathway. Quinolinic acid (150 nmol in 0.5 μ l of phosphate-buffered saline)



(Sigma) was injected stereotactically into the caudate-putamen to produce axon-sparing lesions²⁹. 3 H-FK506 autoradiography was dramatically reduced in the striatum and substantia nigra reticulata ($n=3$). These localization patterns have been replicated 6 to 8 times. Abbreviations: Acb, nucleus accumbens; CA1–CA3, fields of Ammon's horn; CP, caudate putamen; Ctx, cortex; DG, dentate gyrus; Gr, granule cell layer of cerebellum; EP, entopeduncular nucleus; OB, olfactory bulb; OT, olfactory tract; SNr, substantia nigra reticulata; Snp, striatonigral pathway; Tu, olfactory tubercle; T, thalamus; Mol, molecular layers of cerebellum. Scale bar, 2.5 mm.

most evident in the molecular layer. Interestingly, the superior and inferior colliculi are more enriched in calcineurin than FKBP, whereas for mRNA, enrichment is greater for FKBP.

As 300 nM FK506 or cyclosporin A inhibits the phosphatase activity of calcineurin towards a synthetic phosphopeptide substrate³, the immunophilins might regulate phosphorylation physiologically. We have directly observed enhanced phosphorylation of physiological protein substrates in the presence

of very low concentrations of FK506 and cyclosporin A (Fig. 2). We incubated soluble rat-brain extracts with [γ - 32 P]ATP in the presence or absence of FK506, cyclosporin A and *O*-tetradecanoylphorbol 13-acetate (TPA) to activate protein kinase (PKC). As little as 1 nM FK506 enhances the phosphorylation of several protein bands. Cyclosporin A shows a similar activity, although it is 10–100 times less potent than FK506. The most prominent band corresponds to an M_r of $\sim$ 45,000 (45K) when

FIG. 2 FK506 and cyclosporin A mediate phosphorylation of substrate protein *in vitro*. a, Whole rat-brain soluble extracts from 200 mg ml $^{-1}$ wet-weight brain homogenates, were incubated with 50 μ g ml $^{-1}$ phosphatidylserine, 20 μ M [γ - 32 P]ATP in 50 mM HEPES, 1 mM Na EGTA, 2 mM dithiothreitol, pH 7.4, buffer in the absence or presence of 10 μ M free Ca^{2+} , 200 nM TPA and 1, 10 or 100 nM FK506 for 20 min at 25 $^{\circ}$ C in a final vol of 0.2 ml. Reactions were stopped by addition of PAGE sample buffer, proteins were resolved on 3.5–17% linear gradient polyacrylamide gels using the buffers of Laemmli³⁰, gels were dried down and exposed to X-ray film (Kodak XR5 film) and autoradiograms developed. Lane 1, control (no Ca^{2+} , no TPA); lane 2, Ca^{2+} -stimulated phosphorylation; lane 3, Ca^{2+} and TPA-stimulated phosphorylation; lanes 4–6 Ca^{2+} and TPA-stimulated phosphorylation in the presence of 100 nM, 10 nM and 1 nM FK506, respectively. Molecular weight markers are indicated. b, Whole rat-brain soluble extracts were phosphorylated and processed as in a, except that 1 μ M cyclosporin A was included instead of FK506. Lane 1, control (no Ca^{2+} , no TPA); lane 2, Ca^{2+} -stimulated phosphorylation; lane 3, Ca^{2+} and TPA-stimulated phosphorylation; lane 4, Ca^{2+} and TPA-stimulated phosphorylation in the presence of 1 μ M cyclosporin A. Molecular weight markers are indicated.

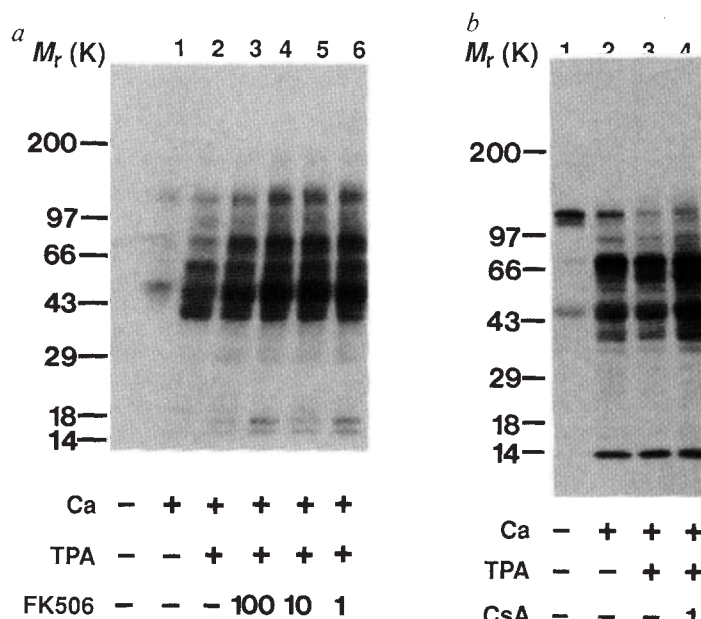


TABLE 1 Distribution of [^3H]FK506 binding sites in brain and peripheral tissues

Tissue	Membrane	Soluble
Cortex	24.5 $\pm$ 2.2	17.5 $\pm$ 1.0
Cerebellum	4.0 $\pm$ 0.3	9.6 $\pm$ 0.7
Striatum	45.8 $\pm$ 3.9	33.8 $\pm$ 2.5
Hippocampus	34.7 $\pm$ 1.5	42.6 $\pm$ 3.2
Brainstem	5.9 $\pm$ 0.6	9.7 $\pm$ 0.4
Hypothalamus	10.3 $\pm$ 0.9	7.5 $\pm$ 0.5
Midbrain	12.8 $\pm$ 1.0	29.9 $\pm$ 1.0
Pituitary	*	11.3 $\pm$ 1.2
Thymus	1.8 $\pm$ 0.2	3.1 $\pm$ 0.4
Spleen	1.4 $\pm$ 0.2	2.0 $\pm$ 0.1
Heart	*	3.6 $\pm$ 0.4
Kidney	0.5 $\pm$ 0.02	3.4 $\pm$ 0.4
Liver	*	3.7 $\pm$ 0.5
Lung	0.7 $\pm$ 0.04	2.6 $\pm$ 0.3

[^3H]FK506 bound is expressed as pmol per mg protein. Regions of rat brain were dissected, homogenized at 100 mg ml $^{-1}$ wet weight in 50 mM Tris-HCl, pH 7.4, 1 mM sodium EDTA, 100 μg ml $^{-1}$ phenylmethylsulfonylfluoride and centrifuged at 40,000g for 20 min at 4°C. Membranes were resuspended in 20 volumes of homogenization buffer, repelleted at 40,000g for 20 min, and resuspended in homogenization buffer to the original volume. Protein concentration was determined by the Coomassie blue dye binding assay²³ using bovine serum albumin as a standard. Binding of 250 pM [^3H]dihydroFK506 (86.5 Ci mmol $^{-1}$; NEN/DuPont) to membrane protein prepared from the indicated rat brain regions was performed in a final volume of 0.5 ml composed of 50 mM Tris-HCl, pH 7.4, 2 mg ml $^{-1}$ BSA buffer (assay buffer) and various concentrations of unlabelled FK506. After 60 min at room temperature incubation, the membrane-bound [^3H]FK506 was recovered by filtration on 0.2% TX100-soaked glass fibre filter strips in a cell harvester (Brandel, Gaithersburg, MD) followed by three 4-ml washes in 50 ml NaCl. The glass filter disk was mixed with 5 ml of Formula 963 scintillation cocktail (NEN/DuPont) and counted in a Beckman scintillation counter. Nonspecific binding in the presence of 1 μM FK506 is subtracted from total binding to yield specific binding. Binding of [^3H]FK506 to peripheral tissue membrane protein was assayed with 300–400 μg membrane protein in assay buffer and 250 pM [^3H]FK506 in a final volume of 0.5 ml. Membrane bound [^3H]FK506 was separated from unbound ligand by centrifugation at 20,000g for 15 min. Nonspecific binding was determined in the presence of 1 μM unlabelled FK506. [^3H]FK506 binding to soluble protein was done as described for cyclosporin A (ref. 24), using 250 pM [^3H]FK506 and 10–50 μg protein in assay buffer to a final volume of 0.4 ml. After 60 min incubation 25°C, 0.35 ml was layered over a 0.8-ml column of LH-20 Sephadex (Pharmacia LKB), pre-equilibrated with assay buffer. The column was further washed with 0.4 ml of assay buffer, the eluates collected, mixed with Formula 963 cocktail and counted. Specific binding was determined by subtracting binding in the presence of 1 μM unlabelled FK506 from total [^3H]FK506 bound. All binding experiments were done in duplicate, with an average range of 7–10%. Similar results have been obtained from three separate determinations of [^3H]FK506 binding levels.

* Binding detected, but at extremely low level (<0.05 pmol mg $^{-1}$).

visualized on 3.5–17% gradient gels, but migrates from 43K to 60K on 7% and 12% polyacrylamide gels, respectively. The mobility changes in different percentage polyacrylamide gels of a PKC-phosphorylated protein of this size are characteristic of GAP43, a prominent calcineurin substrate; indeed, phosphorylation of purified GAP43 by PKC is enhanced by FK506 (Fig. 3) and cyclosporin A (data not shown). We have not identified definitively the protein bands at 27, 60, 80 and 120K whose phosphorylation is augmented by FK506 and cyclosporin A.

To investigate this regulated phosphorylation further, we used intact PC12 cells, a neuronal-like cell line, labelling cellular ATP pools with ^{32}P . At 1 and 10 nM, FK506 enhances phosphorylation of a 27K and a 50K protein, which are also labelled in brain preparations *in vivo* (data not shown).

The extraordinarily high concentrations of FKBP and its mRNA in discrete neuronal populations in the brain suggests that neural functions of the immunophilins may be as important

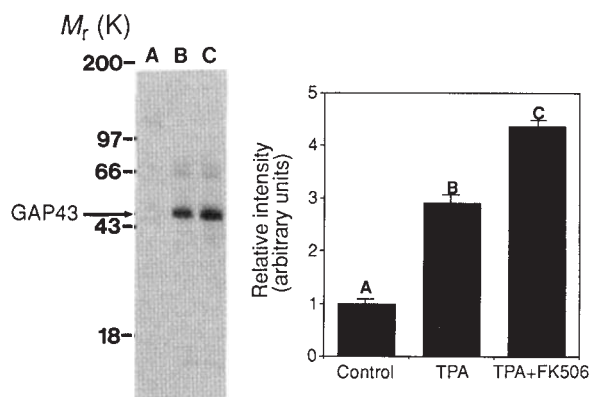


FIG. 3 GAP-43, purified from bovine brain³¹, was incubated with rat brain cytosol (lane A) in the presence of Ca $^{2+}$, TPA (lane B), and Ca $^{2+}$, TPA and 10 nM FK506 (lane C) as described in Fig. 2 legend. The level of phosphorylation of purified GAP-43 was quantitated by densitometry using the Amersham/LOATS RAS3000 imaging system. This experiment has been replicated three times with similar results.

as their role in mediating responses to immunosuppressant drugs. We have not measured brain levels of cyclophilin here, but regional variations in mRNA levels of cyclophilin have been reported, with highest densities in neuronal tissue in the hippocampus and cerebellum⁷.

The immunophilins have peptidyl prolyl isomerase activity which is inhibited by FK506 and cyclosporin A (refs 1, 2, 8–11), but the ability of immunosuppressant drugs to inhibit this activity does not correlate with their immunosuppressant function. The binding of the immunophilins to calcineurin in the presence of their drug ligands may be a more likely mechanism of action as the immune functions regulated by these drugs are invariably Ca $^{2+}$ -dependent processes^{12–15}. The immunophilins regulate exocytosis of secretory granules in mast cells, which is also a Ca $^{2+}$ -dependent process^{16–18}, and regulation of synaptic functions, especially neurotransmitter release, may be a function for immunophilins working with calcineurin in the brain. GAP43, a prominent substrate of calcineurin whose phosphorylation is regulated by immunophilins, is a 'growth-associated protein' for neuronal growth cones^{19–21} but is probably also involved in neurotransmitter release, because antibodies against GAP43 selectively inhibit neurotransmitter release²².

The colocalization of FKBP and calcineurin implies that they interact in normal brain function. As FKBP alters calcineurin activity only when bound to a ligand such as FK506, we propose that endogenous ligands exist for FKBP and cyclophilin that modulate calcineurin activity and the phosphorylation of its substrate proteins. □

Received 10 March; accepted 25 June 1992.

- Schreiber, S. L. *Science* **251**, 283–287 (1991).
- Thomson, A. W. *Immunol. Today* **10**, 6–9 (1989).
- Liu, J. et al. *Cell* **66**, 807–815 (1991).
- Shieh, B. H., Starnes, M. A., Seavello, S., Harris, G. L. & Zuker, C. S. *Nature* **338**, 67–70 (1989).
- Schneuwly, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5390–5394 (1989).
- Maki, N. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5440–5443 (1990).
- Lad, R. P., Smith, M. A. & Hilt, D. C. *Molec. Brain Res.* **9**, 239–244 (1991).
- Takahashi, N., Hayano, T. & Suzuki, M. *Nature* **337**, 473–475 (1989).
- Fischer, G., Wittmann-Liebold, B., Lang, K., Kiehlhaver, T. & Schmid, F. X. *Nature* **337**, 476–478 (1989).
- Siekierka, J. J., Hung, S. H. Y., Poe, M., Lin, C. S. & Sigal, N. *Nature* **341**, 755–757 (1989).
- Harding, M. W., Galat, A., Uehling, D. E. & Schreiber, S. L. *Nature* **341**, 758–760 (1989).
- Dumont, F. J., Staruch, M. J., Koprak, S. L., Melino, M. R. & Sigal, N. H. *J. Immunol.* **144**, 251–258 (1990).
- Goldsmith, M. A. & Weiss, A. *Science* **240**, 1029–1031 (1988).
- Kimball, P. M., Gamaz, N. & Sell, S. *Cell Immunol.* **113**, 107–116 (1988).
- Bierer, B. E., Schreiber, S. L. & Burakoff, S. J. *Eur. J. Immunol.* **21**, 439–445 (1991).
- Hultsch, T., Albers, M. W., Schreiber, S. L. & Hohman, R. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6229–6233 (1991).
- Trenn, G. et al. *J. Immunol.* **142**, 3796–3802 (1989).
- Triggiani, M., Cirillo, R., Lichtenstein, L. M. & Marone, G. *Int. Arch. Allergy appl. Immunol.* **88**, 253–255 (1989).

19. Benowitz, L. I. & Routtenberg, A. *Trends Neurosci.* **10**, 527-531 (1987).
20. Skene, J. H. P. A. *Rev. Neurosci.* **12**, 127-156 (1989).
21. Coggins, P. J. & Zwieters, H. *J. Neurochem.* **56**, 1095-1106 (1991).
22. Dekker, L. V., DeGraan, P. N. E., Oestreicher, A. B., Versteeg, D. H. G. & Gispen, W. H. *Nature* **342**, 74-76 (1989).
23. Bradford, M. M. *Ann. Biochem.* **72**, 248-254 (1976).
24. Handschumacher, R. E., Harding, M. W., Rice, J., Druggie, R. J. & Speicher, D. W. *Science* **226**, 544-546 (1984).
25. Ross, C. A., MacCumber, M. W., Glatt, C. E. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2923-2927 (1990).
26. Standaert, R. F., Galat, A., Verdine, G. L. & Schreiber, S. L. *Nature* **346**, 671-674 (1990).
27. Ito, A. et al. *Biochem. biophys. Res. Commun.* **163**, 1492-1497 (1989).
28. Kuno, T. et al. *Biochem. biophys. Res. Commun.* **165**, 1352-1358 (1989).
29. Schwarz, R., Whetzel, W. D. Jr & Mangao, R. M. *Science* **219**, 316-318 (1983).
30. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
31. Andreason, T. J., Luefje, C. W., Heideman, W. & Storm, D. R. *Biochemistry* **22**, 4618-4624 (1983).

ACKNOWLEDGEMENTS. We thank C. Klee for advice and for antibodies to calcineurin, and the Fujisawa Corporation for unlabelled FK506. Supported by a USPHS grant and a Research Scientist Award to S.H.S., and a grant from the International Life Sciences Institute and a gift from Bristol-Myers-Squibb; T.M.D. is a Pfizer Postdoctoral Fellow and is supported by grants from the American Academy of Neurology, the French Foundation for Alzheimer Research and the Dana Foundation; J.P.S. is supported by a Postdoctoral Fellowship grant and M.F. was supported in part by Fonds pour Formation de Chercheurs et l'Aide à la Recherche (FCAR) of Québec, Canada.

Homodimer formation of retinoid X receptor induced by 9-*cis* retinoic acid

Xiao-kun Zhang, Jürgen Lehmann, Birgit Hoffmann, Marcia I. Dawson*, James Cameron*, Gerhart Graupner, Thomas Hermann, Paul Tran & Magnus Pfahl†

Cancer Center, La Jolla Cancer Research Foundation,
10901 N. Torrey Pines Road, La Jolla, California, USA
* SRI International Life Sciences Division, Menlo Park,
California 94025, USA

RETINOID response pathways are mediated by two classes of receptors, the retinoic acid receptors (RARs)¹⁻⁶ and the retinoid X receptors (RXRs)⁷⁻¹¹. A central question is whether distinct response pathways are regulated by these two classes of receptors. The observation that the stereoisomer 9-*cis*-retinoic acid binds with high affinity to RXRs^{12,13} suggested that this retinoid has a distinct role in controlling RXR activity, but it was almost simultaneously discovered that RXRs function as auxiliary receptors for RARs and related receptors, and are essential for DNA binding and function of those receptors^{9,10,14-17}. Hence, although RARs seem to operate effectively only as heterodimeric RAR/RXR complexes, RXRs themselves apparently function predominantly, if not exclusively, as auxiliary receptors. Here we report that 9-*cis*-retinoic acid induces RXR homodimer formation. Our results demonstrate a new mechanism for retinoid action by which a ligand-induced homodimer mediates a distinct retinoid response pathway.

We investigated the effect of 9-*cis*-retinoic acid on RXR α binding to the palindromic thyroid hormone response element (TREp), an RXR responsive element^{8,14}. In the absence of ligand, RXR alone did not bind effectively to the TREp but required RAR or thyroid hormone receptor (TR) for response element interaction. But in the presence of 9-*cis*-retinoic acid, DNA binding of RXR was dramatically increased (Fig. 1a) and did not require thyroid hormone receptors or RARs. The prominent RXR-specific band observed in the presence of 9-*cis*-retinoic acid migrated more slowly than the TR-RXR complex at a position indistinguishable from that of the RAR-RXR heterodimer complex (Fig. 1a). These data therefore suggested that 9-*cis*-retinoic acid induced RXR homodimer binding to the TREp. Interestingly, the retinoid did not prevent heterodimer interaction because both TR-RXR heterodimers and RXR

homodimers were observed when thyroid hormone receptor was mixed with RXR in the presence of 9-*cis*-retinoic acid. To determine whether RXR homodimers could also form in the presence of RARs, we used RAR or RXR protein derivatives that carried the Flag epitope^{14,18}. When Flag-RXR was incubated with RAR in the presence of 9-*cis*-retinoic acid and the anti-Flag antibody, a complete supershift was observed. But when Flag-RAR was used under the same conditions, there was only a partial supershift, in support of the presence of RXR homodimers that could not react with the antibody (Fig. 1b).

Remarkably, all-*trans*-retinoic acid (10⁻⁶ M) also induced RXR α homodimer binding to some degree, whereas thyroid hormone T3 did not. RAR homodimer DNA binding was not induced by 9-*cis*-retinoic acid (Fig. 1a). Formation of the 9-*cis*-retinoic acid-induced complex was dependent on the concentration of RXR protein and was strongly cooperative (Fig. 1c). RXR complex formation was enhanced by 9-*cis*-retinoic acid at all receptor concentrations tested, being already effective at 10⁻⁹ M, with optimal binding at 10⁻⁸ M (Fig. 1d).

We next tested RXR interaction with several natural and synthetic response elements (Fig. 2a). In the case of the ApoAI-retinoic acid response element (RARE)¹⁸, a strong RXR complex was observed in the presence of 9-*cis*-retinoic acid and a weaker one with retinoic acid (10⁻⁶ M). The RAR-RXR heterodimer also bound effectively to this response element (Fig. 2b). Homodimer complexes were also induced by 9-*cis*-retinoic acid with the CRBP-II-RARE (ref. 19) and the β -RARE (refs. 20, 21; Fig. 2c) elements that contain 1-base-pair (bp) or 5-bp spacers, respectively. In this latter case, however, the RXR homodimer band was considerably weaker than the RAR-RXR heterodimer band. Interestingly, another natural RARE derived from the rat CRBPI promoter (M. Husmann and M. P., unpublished results), which is similar to the ApoAI-RARE in that it contains a 2-bp spacer, did not bind RXR at all in the presence of 9-*cis*-retinoic acid (Fig. 2c), indicating that the actual sequence of the motif is critical for RXR homodimer binding and not simply the spacing separating the two half-sites. Similarly, the DR5-RARE (ref. 22), a perfect repeat element derived from the β -RARE, did not interact with RXR in the presence of 9-*cis*-retinoic acid, although it interacted strongly with the RAR-RXR heterodimer (Fig. 2c).

None of the T3 response elements shown here (Fig. 1a), the rat α -myosin heavy chain TRE (ref. 23), the rat malic enzyme TRE (ref. 24) and the perfect repeat DR4 (ref. 22), bound the RXR in the presence of 9-*cis*-retinoic acid, but all three response elements interacted effectively with TR-RXR heterodimers. Similarly, the perfect palindromic oestrogen response element (ERE; ref. 25) did not measurably bind RXR homodimers (data not shown).

To determine whether 9-*cis*-retinoic acid can induce RXR homodimer formation in solution, we again used the Flag-RXR derivative. When we mixed Flag-RXR with *in vitro*-labelled ³⁵S-RXR protein in the presence of 9-*cis*-retinoic acid, the labelled RXR could be coprecipitated by anti-Flag antibody but not by nonspecific serum (Fig. 3). Coprecipitation efficiency was slightly increased in the presence of the ApoAI-RARE but not in the presence of the α -myosin heavy chain TRE. In addition, incubation with the protein crosslinker DSP (dithiobis-succinimidyl propionate) enhanced coprecipitation of labelled RXR. Specific coprecipitation in all cases, was observed only in the presence of 9-*cis*-retinoic acid. Thus 9-*cis*-retinoic acid induces RXR homodimer formation in solution in the absence of DNA.

In transient transfection assays with a TREp-containing reporter, we found activation by RXR α was strong in the presence of 9-*cis*-retinoic acid and weak when retinoic acid was added. Activation could be enhanced by cotransfection of RAR α . In this case, however, retinoic acid was also an effective activator, although not as efficient as 9-*cis*-retinoic acid (Fig. 4a). The β -RARE (Fig. 4b) was highly activated by endogenous CV-1

† To whom correspondence should be addressed.

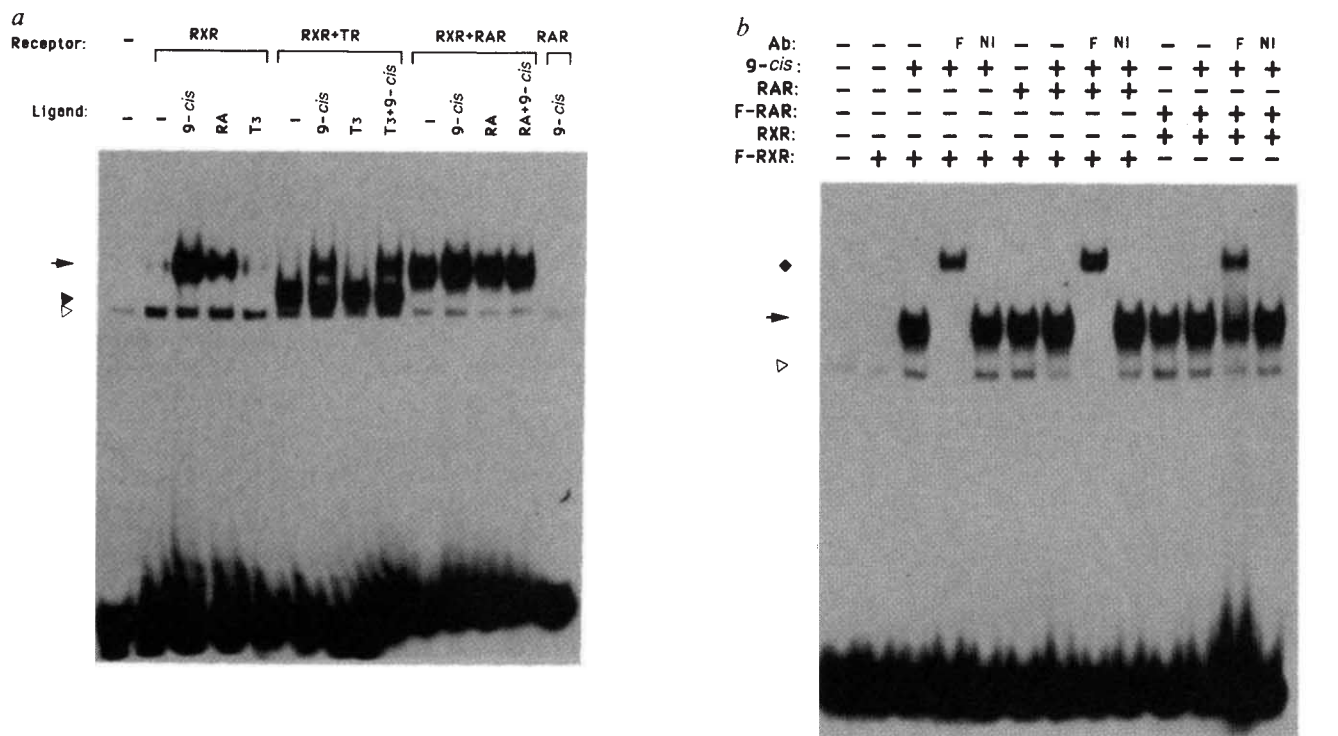


FIG. 1 9-*cis* retinoic acid induces RXR homodimer binding. **a**, *In vitro* synthesized RXR α was incubated either with or without (-) the indicated hormones (10^{-7} M 9-*cis*-retinoic acid (RA); 10^{-6} M RA; 10^{-6} M T3) in the presence or absence of *in vitro*-synthesized TR α or RAR β for 30 min with 32 P-labelled TREp and analysed by gel retardation. Lane 1 shows the nonspecific binding of unprogrammed reticulocyte lysate. The open triangle indicates a nonspecific band observed with unprogrammed reticulocyte lysate. Solid triangle indicates the specific TR α -RXR α heterodimer binding. Arrow

indicates specific RXR α homodimer binding. The RAR β -RXR α heterodimer migrates at the same position as the RXR α homodimer. For comparison, the effect of 9-*cis*-RA on RAR β DNA binding is also shown. **b**, To show that 9-*cis*-RA-induced DNA binding complex contains RXR α protein and to distinguish between RXR homodimer and RAR-RXR heterodimer, Flag-RXR α (F-RXR) and Flag-RAR (F-RAR), which can be recognized by a specific anti-Flag monoclonal antibody, were constructed. *In vitro*-synthesized F-RXR α and other receptors, as indicated, were incubated either with (+) or without (-) 10^{-7} M 9-*cis*-RA in the presence of either antibody (Ab) specific against Flag (F) or nonspecific preimmune serum (NI). The complexes formed were then analysed by gel retardation assay using 32 P-labelled TREp as a probe. Lane 1 represents the nonspecific binding of unprogrammed reticulocyte lysate (open triangle). Arrow indicates the specific F-RXR α homodimer and RAR-RXR heterodimer binding. Diamond indicates the anti-Flag antibody upshifted receptor complex. **c**, Formation of RXR-TREp complexes at different receptor concentrations. Open triangle indicates nonspecific binding. Arrows indicate the RXR complex. **d**, RXR homodimer binding occurs at low 9-*cis*-RA concentrations. Equal amounts of *in vitro*-synthesized RXR protein were incubated with the indicated concentrations (molar) of 9-*cis*-RA and labelled TREp. Open triangles indicate nonspecific binding of unprogrammed reticulocyte lysate. Arrow indicates RXR complex.

METHODS. The cloning of the receptor complementary DNAs for RXR α , RAR β ,

or TR α into the pBluescript (Stratagene) have been described¹⁴. Flag-RXR α and Flag-RAR were constructed as before^{27,28} by ligation of a double-stranded oligonucleotide containing an ATG codon and a DNA sequence encoding Flag (Arg-Tyr-Lys-Asp-Asp-Asp-Lys) to the N-terminus of RXR α or RAR γ . The hybrid gene was then cloned into pBluescript. Synthesis of receptor proteins using an *in vitro* transcription/translation system and gel retardation assays, as well as the double-stranded TREp²⁹, were all as described²⁷. 9-*cis*-RA (m.p. 184–187 °C) was prepared from 9-*cis*-retinal by a two-step sequence of MnO₂/NaCN oxidation in the presence of acetic acid/methanol to give the methyl ester (69%) followed by hydrolysis (80%) in 0.5 M KOH in 25% aqueous methanol and crystallization (methanol); HPLC (Novapak C18, 32% CH₃CN, 27% methanol, 16% isopropanol, 24% H₂O, 1% acetic acid 1.9 ml min⁻¹, 260 nm), retention time 15.4 min (100%). To analyse the effect of hormones, receptor proteins were incubated with appropriate concentrations of hormone at room temperature for 30 min before assaying for DNA binding. When anti-Flag antibody (from M. Leahy) was used, 1 μ l antiserum was incubated with receptor protein for another 30 min at room temperature before the DNA-binding assay.

<i>a</i> TRE _{pal}	gatcTCAGGTCATGACCTGAgatc ctagAGTCCAGTACTGGACTctag
ApoA1-RARE	gatcAGGGCAGGGTCAAGGGTTCAGTgatc ctagTCCCGTCCCAAGTCCCAAGTCActag
CRBP11-RARE	gatcCAGGTCACAGGTCACAGGTCACAGTTCAGatc ctagGTCCAGTGTCCAGTGTCCAGTGTCAAGTTCtag
β RARE	gatctGTAGGGTTCACGAAAGTTCAGTCagatc ctagaCATCCCAAGTGGCTTCAAGTGAgtctag
CRBP1-RARE	gatccAGGTCAAAAGTCAGgatc ctaggTCCAGTTTTCAGTCctag
MHC-TRE	gatcCTGGAGGTGACAGGAGACAGCgatc ctagGACCTCCAGTGTCTCTCTGCGctag
ME-TRE	gatcCAGGACGTGGGGTTAGGGGAGGACAGTGGgatc ctagGTCTGCAACCCCAATCCCTCCTGTACCCtag
DR-4	gatcTCAGGTCATCTCAGGTCagatc ctagAGTCCAGTAGAGTCCAGTctag
DR-5	gatcTCAGGTCATCCTCAGGTCagatc ctagAGTCCAGTAGGAGTCCAGTctag
ERE	gatcTCAGGTCATGTGACCTGAgatc ctagAGTCCAGTGACATGGACTctag

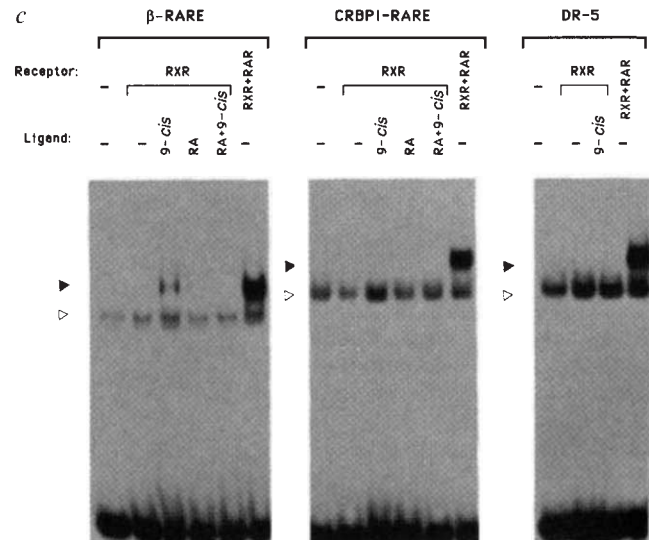
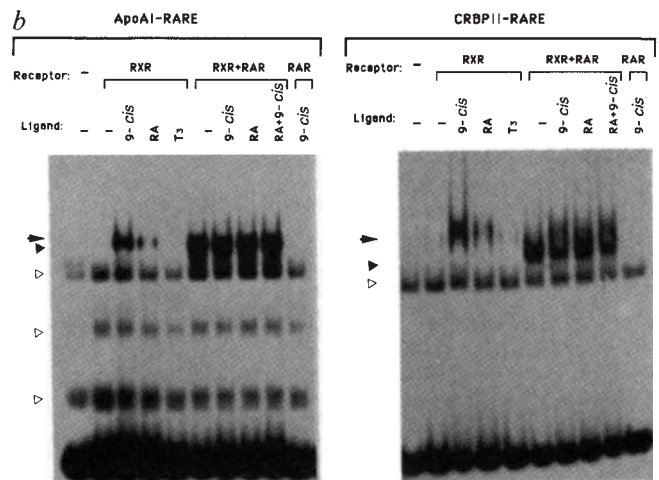


FIG. 2 *9-cis*-retinoic acid induces RXR homodimer binding on RXR specific response elements. *a*, Receptor response elements used in this study. Oligonucleotides were synthesized with appropriate restriction sites at both ends as indicated by lower-case letters. Sequences that are closely related to the AGG/TTCA motif are indicated by arrows (for references, see Methods). *b*, Effect of *9-cis*-RA and other ligands on RXR homodimer and heterodimer binding to the ApoA1-RARE or the CRBP11-RARE. *c*, Effect of *9-cis*-RA and other ligands on RXR homo- and heterodimer binding to RA-responsive elements. Nonspecific binding of unprogrammed reticulocyte lysate is indicated by open triangles. Solid triangles indicate the heterodimer complexes. Arrows indicate the RXR homodimer complexes.

METHODS. Gel retardation assays using *in vitro*-synthesized receptor protein were done as described^{14,27}. ³²P-Labelled double-stranded oligonucleotides shown in *a* were used as probes. ApoA1-RARE is a direct repeat response element with 2-bp space¹⁸; CRP11-RARE, a direct repeat RXR-specific response element with 1-bp spacer¹⁹; β -RARE, a direct repeat with a 5-bp spacer^{20,21}; CRBP1-RARE, a direct repeat RA-specific response element present in rat CRBP1 promoter (M. Husmann and M.P., unpublished results); MHC-TRE, a direct repeat T3-specific response element present in the rat gene for α -myosin heavy chain²³; ME-TRE, a direct repeat T3-specific response element present in the rat gene for malic enzyme²⁴; DR-4 is an idealized direct repeat T3-specific response element with 4-bp spacer²²; DR-5 is an idealized direct repeat RA-specific response element with 5-bp spacer²²; ERE is a perfect palindromic oestrogen response element²⁵ that differs from TRE by a 3-nucleotide spacer²⁹.

cell receptors, in agreement with previous observations^{21,26}, so we could not detect any further activation by low concentrations of cotransfected receptors. Interestingly, however, *9-cis*-retinoic acid gave a more potent activation at 10^{-7} M than retinoic acid. Thus the endogenous CV-1 cell retinoid receptors are responsive to *9-cis*-retinoic acid.

The ApoA1 element-containing reporter was very effectively activated by RXR α in the presence of *9-cis*-retinoic acid (Fig. 4c), whereas retinoic acid showed no RXR α -specific induction. Maximal activation was seen when both RXR α and RAR α were cotransfected. Under these conditions retinoic acid was also a strong inducer. In contrast, the CRBP1 element, which did not bind RXR homodimer, was not activated by RXR and *9-cis*-retinoic acid either (Fig. 4d), whereas RAR α in the presence of *9-cis*-retinoic acid led to significant induction. The heterodimer RAR α in the presence of *9-cis*-retinoic acid led to significant induction. The heterodimer RAR α -RXR α most

efficiently activated this response element in the presence of *9-cis*-retinoic acid. Adding both ligands (*9-cis*- and all-*trans*-retinoic acid) gave no further induction of any of the RAREs tested (data not shown). Not unexpectedly, there was no induction by RXR α and *9-cis*-retinoic acid with the reporters containing α -myosin heavy chain or malic enzyme TREs or the ERE (data not shown). The transactivation results thus correlate directly with results from *in vitro* DNA-binding studies, in that effective activation by RXR α in the presence of *9-cis*-retinoic acid is only observed on response elements that strongly interact with the *9-cis*-retinoic acid induced RXR α homodimer.

Our analyses show that the natural vitamin A derivative *9-cis*-retinoic acid, in contrast to all-*trans*-retinoic acid, leads to effective RXR homodimer formation and that these homodimers bind and activate several RAREs but not natural TREs or the ERE. The response element specificity of RXR homodimers seems to be distinct from heterodimers: the CRBP1

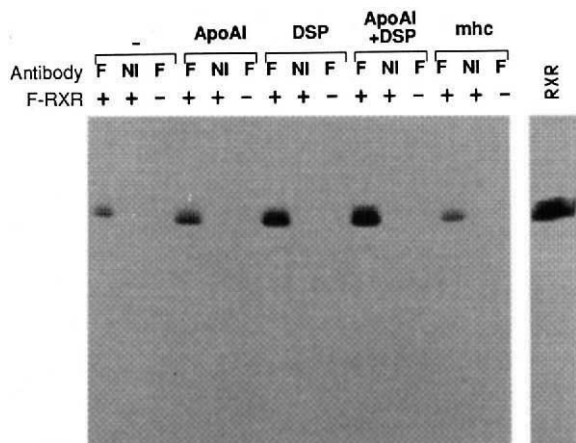


FIG. 3 RXR homodimerization occurs in solution. ^{35}S -labelled *in vitro*-synthesized RXR α protein was incubated with partially purified bacterially expressed Flag-RXR (F-RXR) (+) or similarly prepared glutathione S-transferase control protein (-) either in the presence or absence of response elements or the protein crosslinker DSP as indicated. After incubation, either anti-Flag antibody (F) or nonspecific preimmune serum (NI) was added. The immune complexes were washed in the presence of 10^{-7} M 9-*cis*-RA, boiled in SDS sample buffer and separated on a 10% SDS-polyacrylamide gel. The ^{35}S -labelled *in vitro*-synthesized RXR α protein is shown on the right. METHODS. Flag-RXR α was cloned in-frame in the expression vector pGex2T (Pharmacia) and was expressed as a glutathione S-transferase fusion protein in bacteria following the manufacturer's protocol. The protein was purified (~90%) on a prepacked glutathione-Sepharose-4B column (Pharmacia) and tested for its function by gel retardation assays and western blotting using anti-Flag antibody. The immunoprecipitation assay was as described¹⁴. Briefly, 10 μl ^{35}S -labelled *in vitro*-synthesized RXR α protein were incubated with 5 μl (~0.1 μg) partially purified bacterially expressed Flag-RXR α fusion protein or similarly prepared glutathione S-transferase control protein in 100 μl buffer containing 10^{-7} M 9-*cis*-RA, 50 mM KCl and 10% glycerol for 30 min at room temperature. When assayed in the presence of chemical crosslinker or oligonucleotides, we added 2 μl 100 mM dithiobis(succinimidylpropionate) (DSP) dissolved in dimethyl sulphoxide or 10 ng oligonucleotide and continued the incubation at room temperature for 15 min. Reaction mixtures were then incubated with 1 μl anti-Flag antibody or nonspecific preimmune serum for 2 h on ice. Immune complexes were precipitated by adding 50 μl protein A-Sepharose slurry and mixing continuously in the cold room for 1 h. The immune complexes were washed extensively with cold NET-N buffer (20 mM Tris, pH 8.0, 100 mM NaCl, 1 mM dithiothreitol, 0.5% NP40) containing 10^{-7} M 9-*cis*-RA, boiled in SDS sample buffer, and resolved by SDS-PAGE. The gel was fixed, dried and visualized by autoradiography.

response element did not interact with RXR homodimers and the CRBP2 response element effectively bound 9-*cis*-retinoic acid-induced RXR α homodimers. Although the CRBP2 response element also binds the RAR-RXR heterodimers (Fig. 2), the heterodimer appears to have a repressor function¹⁹. Our results also indicate that spacing of the repeat motifs cannot itself distinguish between responsiveness of RAREs to either heterodimers or homodimers. Thus 9-*cis*-retinoic acid can change the equilibrium between monomeric, homodimeric and heterodimeric receptors, generating new response pathways. Overall, our data demonstrate the central roles of the retinoid X receptors in acting as auxiliary receptors for several classes of hormone receptors and in functioning independently as homodimers in the presence of 9-*cis*-retinoic acid. The equilibrium between homodimers and heterodimers is likely to control not only distinct retinoid pathways, but also the ligand responses of receptors that form heterodimers with RXRs. This, combined with the possible development of RXR-selective ligands, may provide a suitable basis for new therapeutic approaches targeting

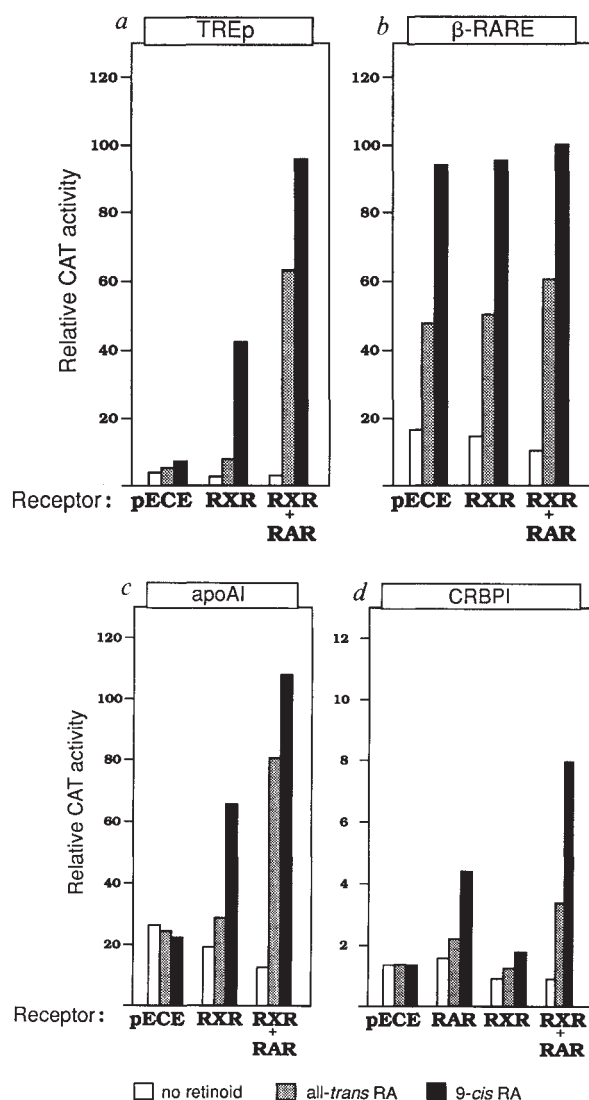


FIG. 4 Transcriptional activation of RXR α -, RAR α - and RXR-containing heterodimers by 9-*cis*-RA. CV-1 cells were cotransfected with 100 ng reporter plasmids: a, TREp-tk-CAT; b, β-RARE-tk-CAT; c, ApoAI-RARE-tk-CAT; and d, CRBPI-RARE-tk-CAT and 5 ng empty pECE expression vector, pECE-RXR α , pECE-RAR α , or combination of both as indicated. Transfected cells were treated with no hormone (open bars), 10^{-7} M RA (shaded bars) or 10^{-7} M 9-*cis*-RA (filled bars). Results of a representative experiment ($n=2$) are shown. tk, Thymidine kinase; CAT, chloramphenicol acetyltransferase. METHODS. CV-1 cells were transiently transfected using a modified calcium phosphate precipitation procedure as described previously²⁶. The reporter constructs and expression vectors have been described previously^{14,27,30}. CAT activity was normalized for transfection efficiency by measuring the enzymatic activity derived from the cotransfected β-galactosidase expression plasmid (pCH110; Pharmacia).

diseases known to respond only to hyperphysiological doses of retinoic acid. As only the RXR-responsive pathway will be affected, the severe side effects of the retinoic acid response may be avoided. □

Received 29 April; accepted 23 June 1992.

- Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444-450 (1987).
- Giguere, V., Ong, E. S., Segli, P. & Evans, R. M. *Nature* **330**, 624-629 (1987).
- Benbrook, D., Lernhardt, E. & Pfahl, M. *Nature* **333**, 669-672 (1988).
- Brand, N. *et al.* *Nature* **332**, 850-853 (1988).
- Krust, A., Kastner, P. H., Petkovich, M., Zelen, A. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5310-5314 (1989).
- Giguere, V. *et al.* *Molec. cell. Biol.* **10**, 2335-2340 (1990).
- Hamada, K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 8289-8293 (1989).

8. Mangelsdorf, D. J., Ong, E. S., Dyck, J. A. & Evans, R. M. *Nature* **345**, 224–229 (1990).
9. Yu, V. C. et al. *Cell* **67**, 1251–1266 (1991).
10. Leid, M. et al. *Cell* **68**, 377–395 (1992).
11. Mangelsdorf, D. J. et al. *Genes Dev.* **6**, 329–344 (1992).
12. Heyman, R. et al. *Cell* **68**, 397–406 (1992).
13. Levin, A. A. et al. *Nature* **355**, 359–361 (1992).
14. Zhang, X.-K., Hoffmann, B., Tran, P., Graupner, G. & Pfahl, M. *Nature* **355**, 441–446 (1992).
15. Kiewer, S. A., Umesono, K., Mangelsdorf, D. J. & Evans, R. M. *Nature* **355**, 446–449 (1992).
16. Marks, M. S. et al. *EMBO J.* **11**, 1419–1435 (1992).
17. Bugge, T. H., Pohl, J., Lonnay, O. & Stunnenberg, H. G. *EMBO J.* **11**, 1409–1418 (1992).
18. Rottman, J. N., Widom, R. L., Nadal-Ginard, B., Mahdavi, V. & Karathanasis, S. K. *Molec. cell. Biol.* **11**, 3814–3820 (1991).
19. Mangelsdorf, D. J. et al. *Cell* **66**, 555–561 (1991).
20. de Thè, H., Vivanco-Ruiz, M. M., Tiollais, P., Stunnenberg, H. & Dejean, A. *Nature* **343**, 177–180 (1990).
21. Hoffmann, B. et al. *Molec. Endocr.* **4**, 1734–1743 (1990).
22. Umesono, K., Murakami, K. K., Thompson, C. C. & Evans, R. M. *Cell* **65**, 1255–1266 (1991).
23. Flink, I. L. & Morkin, E. J. *biochem. J.* **265**, 11233–11237 (1990).
24. Desvergne, B., Petty, K. J. & Nikodem, V. M. *J. Biol. Chem.* **266**, 1008–1013 (1991).
25. Klein-Hitpass, L., Schorpp, M., Wagner, U. & Ryffel, G. U. *Cell* **46**, 1053–1061 (1986).
26. Husmann, M. et al. *Molec. cell. Biol.* **11**, 4097–4103 (1991).
27. Zhang, X.-K., Tran, P. & Pfahl, M. *Molec. Endocr.* **5**, 1909–1920 (1991).
28. Hermann, T. et al. *Cell Reg.* **2**, 565–574 (1991).
29. Glass, C. K., Holloway, J. M., Devary, O. V. & Rosenfeld, M. G. *Cell* **54**, 313–323 (1988).
30. Hermann, T., Hoffmann, B., Zhang, X.-K., Tran, P. & Pfahl, M. *Molec. Endocr.* (in press).

ACKNOWLEDGEMENTS. We thank S. Smith-Ortega for preparing the manuscript, S. Delgado for artwork, and P. Haefner, M. Pfahl and Dun Ha for technical help. X.-K. Z. was supported by a postdoctoral fellowship from the Tobacco-Related Disease Research Programme of the University of California. This work was supported by grants from the NIH (M.D. and M.P.).

The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*

Ying Zhang*, Beate Heym†‡, Bryan Allen§, Douglas Young* & Stewart Cole†

* MRC Tuberculosis and Related Infections Unit, and § Department of Bacteriology, Royal Postgraduate Medical School, Hammersmith Hospital, London W12 0HS, UK

† Laboratoire de Genetique Molculaire Bacterienne, Institut Pasteur, 75724 Paris 15, France

‡ Service de Bacteriologie-Virologie, CHU Pitie-Salpetriere, 75634 Paris 13, France

TUBERCULOSIS is responsible for one in four of all avoidable adult deaths in developing countries¹. Increased frequency and accelerated fatality of the disease among individuals infected with human immunodeficiency virus has raised worldwide concern that control programmes may be inadequate², and the emergence of multidrug-resistant strains of *Mycobacterium tuberculosis* has resulted in several recent fatal outbreaks in the United States³. Isonicotinic acid hydrazide (isoniazid, INH) forms the core of antituberculosis regimens; however, clinical isolates that are resistant to INH show reduced catalase activity and a relative lack of virulence in guinea-pigs^{4–7}. Here we use mycobacterial genetics^{8,9} to study the molecular basis of INH resistance. A single *M. tuberculosis* gene, *katG*, encoding both catalase and peroxidase, restored sensitivity to INH in a resistant mutant of *Mycobacterium smegmatis*, and conferred INH susceptibility in some strains of *Escherichia coli*. Deletion of *katG* from the chromosome was associated with INH resistance in two patient isolates of *M. tuberculosis*.

Most strains of *M. tuberculosis* are highly susceptible to INH (minimum inhibitory concentration, $IC_{min} < 0.02 \mu g ml^{-1}$). Other species of mycobacteria show a range of sensitivities to INH, saprophytic mycobacteria such as *M. smegmatis*, for example, being susceptible to only high concentrations (IC_{min} , $32 \mu g ml^{-1}$) (Fig. 1). In order to identify the *M. tuberculosis* genes involved in INH resistance, a mutant strain of *M. smegmatis*, BH1, which is able to grow at $500 \mu g ml^{-1}$ INH, was isolated¹⁰ and transformed with a set of shuttle cosmid clones containing inserts of about 30 kilobases (kb) with a representative coverage of genomic DNA from *M. tuberculosis* H37Rv. Transformants were tested for growth on plates containing

$32 \mu g ml^{-1}$ INH and one of them, pBH4, was found to confer hypersensitivity to INH ($IC_{min} 8 \mu g ml^{-1}$) (Fig. 1). Minimum inhibitory concentrations for other antituberculosis drugs were unchanged. BH1 has reduced catalase activity by comparison with the parent strain, MC²155 (refs 9, 10), but introduction of pBH4 led to 2–3-fold overproduction of this enzyme (Fig. 1).

Analysis of the catalase activity in *M. tuberculosis* indicated that it resembled the hydroperoxidase I (KatG) enzyme of *E. coli* in being heat-labile and having an associated peroxidase activity¹¹. As a strategy for cloning the relevant gene from *M. tuberculosis*, a series of oligonucleotide probes was designed on the basis of amino-acid sequences conserved between hydroperoxidase I enzymes of *E. coli* and *Bacillus stearothermophilus*^{12,13}. In Southern blot experiments one of these probes, corresponding to amino-acid residues 99–111 of the *E. coli* enzyme identified a 4.5-kb fragment generated by *KpnI* digestion of genomic DNA from *M. tuberculosis* H37Rv. A partial library was prepared from appropriately sized fragments of *KpnI*-digested DNA in the vector pUC19 and a clone containing the corresponding fragment, pYZ55, was isolated by colony hybridization. Transformation of *E. coli* with pYZ55 resulted in expression of peroxidase and catalase activities with electrophoretic mobilities identical to those in extracts of *M. tuberculosis*, and nucleotide sequence analysis identified an open reading frame encoding a protein with marked homology to *E. coli* hydroperoxidase I (Fig. 2).

Restriction enzyme mapping and Southern blot hybridization revealed that pBH4, the cosmid responsible for restoration of INH susceptibility, and pYZ55 contained the same 4.5-kb fragment (Fig. 1). Subsequent transformation of BH1 with this fragment in a mycobacterial shuttle plasmid was shown to confer INH sensitivity; this activity was encoded by a 2.9-kb *EcoRV*-*KpnI* subfragment with coding capacity for the catalase-peroxidase alone (Fig. 1). Inactivation of *katG* by various 3' deletions resulted in loss of both enzyme activity and ability to confer INH susceptibility (Figs 1 and 2).

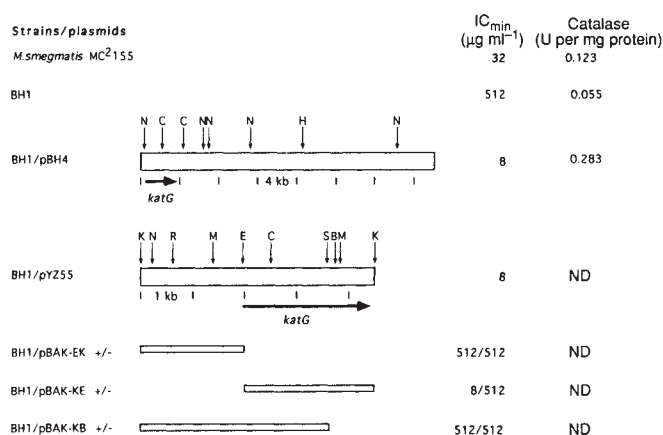


FIG. 1 Restriction map of the DNA insert from pBH4 is shown with that of the insert from pYZ55—a plasmid containing *katG* of *M. tuberculosis* H37Rv, isolated on the basis of hybridization with an oligonucleotide probe (5'-TTCATCCGATGGCCTGGCAGCGCGGGCACCCTACCGC-3') designed to match the amino-acid sequence from a conserved region of *E. coli* hydroperoxidase I. Restriction sites for the following enzymes are indicated: B, *Bam*HI; C, *Cla*I; E, *EcoRV*; H, *Hind*III; K, *Kpn*I; M, *Sma*I; N, *Not*I; R, *Eco*RI; S, *Sac*I. The INH-resistant *M. smegmatis* strain, BH1 (ref. 10); a derivative of strain MC²155 (ref. 9) was transformed with a pool of *M. tuberculosis* H37Rv shuttle cosmids (provided by W. R. Jacobs, New York) and individual clones were scored for INH susceptibility. Cosmid pBH4 consistently conferred drug susceptibility and the transformant overproduced catalase (assayed as described¹⁰). Transformation of BH1 with a mycobacterial shuttle plasmid, pBAK14 (ref. 18), which contained the 4.5-kb insert from pYZ55, likewise conferred INH susceptibility. IC_{min} values are also shown for BH1 transformed with subfragments derived from pYZ55 and inserted into pBAK14 in one (+) or other (–) orientation. The *katG* gene and the ability to confer INH susceptibility both mapped to a 2.9-kb *EcoRV*-*KpnI* fragment (pBAK-KE+). ND, not determined.

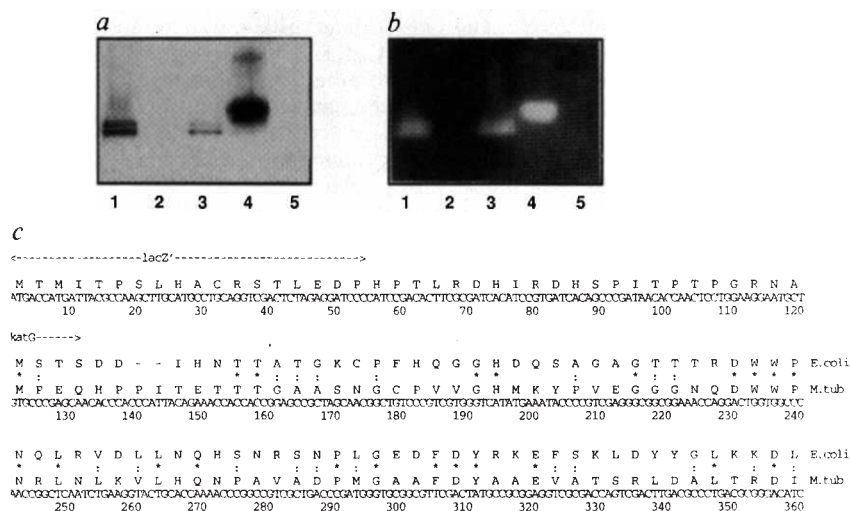


FIG. 2 Extracts from *M. tuberculosis* H37Rv and from *E. coli* strains transformed with a variety of plasmid constructs were prepared for activity gel analysis as described¹⁸. Non-denaturing gels containing 8% polyacrylamide were stained for a, catalase and b, peroxidase activities as described¹⁹. Lane 1, *M. tuberculosis* H37Rv; lane 2, *E. coli* UM2 (*katE*, *katG*; ref. 14); lane 3, *E. coli* UM2/pYZ55; lane 4, *E. coli* UM2/pYZ56 (the 2.9-kb *EcoRV*-*KpnI* fragment in pUC19, corresponding to pBAK-KE+ in Fig. 1); lane 5, *E. coli* UM2/pYZ57 (pYZ55 with a *Bam*HI-*Kpn*I deletion, corresponding to pBAK-KB+ in Fig. 1). *M. tuberculosis* catalase and peroxidase activities migrated as two bands under these conditions (lane 1); the same pattern was seen for the recombinant enzyme expressed by pYZ55 (lane 3). pYZ56 (lane 4) expresses a protein of increased *M.*, owing to a fusion between *katG* and *lacZ'* from the vector shown in c. c, Partial sequence alignment with *E. coli* hydroperoxidase I (the complete sequence of the gene will be communicated elsewhere).

Escherichia coli is completely insensitive to INH. To assess the possibility that expression of the *M. tuberculosis* *katG* gene could confer INH sensitivity in *E. coli*, various strains were transformed with constructs expressing different levels of the enzyme (Fig. 2). No effect was seen in wild-type *E. coli*, but overexpression of the *M. tuberculosis* gene in two catalase-negative strains (UM2 and UM255, both *katG*, *katE*; ref. 14) resulted in susceptibility to high INH concentrations (Fig. 3). Expression of the enzyme at a lower level was less effective in conferring drug susceptibility (Fig. 3).

Many INH-resistant isolates of *M. tuberculosis* have decreased catalase activity, with the most highly resistant isolates ($IC_{min} > 50 \mu g ml^{-1}$) being completely catalase-negative⁴⁻⁷. To examine this phenomenon at the genetic level, chromosomal DNA from a panel of INH-sensitive and INH-resistant strains of *M. tuberculosis* was probed by Southern hybridization using the 4.5-kb *KpnI* fragment containing the catalase-peroxidase gene (Fig. 4). In two highly resistant isolates (strain B1453, and strain 24;

Fig. 4a, lanes 3 and 4) the catalase probe produced no signal at all, whereas in the remaining strains (including a low-resistance isolate, strain 12; IC_{min} , $1.6 \mu g ml^{-1}$), a 4.5-kb *KpnI* fragment identical to that in strain H37Rv was observed. Reprobing of the same blot with the superoxide dismutase gene gave a strong signal with both of the INH-resistant isolates (Fig. 4b), showing that the lack of hybridization with the catalase probe was not due to insufficient DNA. We have previously described an additional genetic change in strain B1453 which results in a restriction fragment length polymorphism like that seen with the superoxide dismutase gene probe (Fig. 4b, lane 3)¹⁵. From a panel of eight INH-resistant isolates, with IC_{min} of 1.6 to $> 50 \mu g ml^{-1}$, *katG* deletion was observed in two out of three highly-resistant strains ($IC_{min} > 50 \mu g ml^{-1}$). We conclude that, in a subset of INH-resistant patient isolates, the loss of catalase activity is due to deletion of the catalase gene.

Mycolic acid biosynthesis, NAD^+ and pyridoxal phosphate metabolism have all been proposed as possible targets for isoniazid^{6,7,16}. Our evidence demonstrates a key role for the catalase-peroxidase enzyme in the action of INH. Deletion of the catalase-peroxidase gene in INH-resistant isolates of *M. tuberculosis* may be an event that is coincident with the deletion of an adjacent undefined gene, but the fact that the catalase-peroxidase gene alone can confer sensitivity to INH in the *M. smegmatis* mutant and in catalase-negative *E. coli* shows that this enzyme must be important in the action of INH. We anticipate that transformation of INH-resistant *M. tuberculosis* isolates with the catalase-peroxidase gene will also restore sensitivity to INH, but transformation experiments with these strains using standard shuttle vectors have so far been unsuccessful. It has been proposed that the catalase-peroxidase enzyme may be important in converting INH to a metabolically active form in the cell^{7,11}, or in the INH-dependent generation of reactive oxygen radicals¹⁷. Availability of the cloned *katG* gene will facilitate the testing of such hypotheses.

Gene deletion represents an unexpected and unusual mechanism for development of drug resistance. Effective drugs must be active against cell components that are essential for bacterial viability, and resistance is generally conferred either by an altered structure of the drug target, or by acquisition of an efficient drug degradation system or permeability barrier. It is likely that in *M. tuberculosis* other forms of INH resistance may also occur. Inactivation of the catalase-peroxidase gene by movement of an insertion element, or point mutations, are attractive theoretical possibilities, and screening of extended panels of INH-resistant isolates of *M. tuberculosis* will be required to assess the relative frequency of gene deletion compared with other potential mechanisms of INH resistance. The

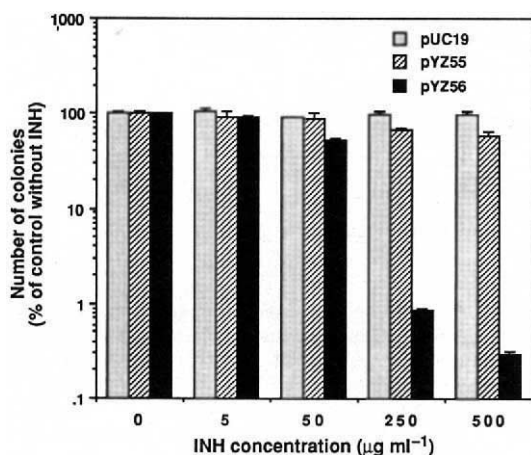
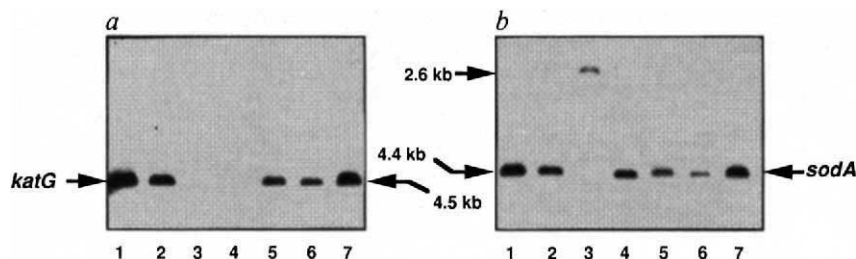


FIG. 3 An *E. coli* strain with mutations in both *katG* and *katE* (UM2; ref. 14) was transformed with pUC19 vector alone (stippled bars), pYZ55 expressing *M. tuberculosis* *katG* (hatched bars) and pYZ56 with high-level expression of *M. tuberculosis* *katG* (solid bars). Overnight cultures in Luria-Bertani broth supplemented with appropriate antibiotics were plated out in the presence of varying concentrations of INH and colony-forming units were assessed. Results of a representative experiment are shown with error bars indicating the standard deviation in triplicate samples. Overexpression of *M. tuberculosis* *katG* similarly conferred susceptibility to high concentrations of INH in *E. coli* UM255 (*katG*, *katE*; ref. 14), but had no effect on catalase-positive strains such as *E. coli* TG1.

FIG. 4 Southern blots prepared using genomic DNA from different *M. tuberculosis* strains, digested with *KpnI*, were probed with *a*, *katG* (the 4.5-kb *KpnI*-*KpnI* fragment), and *b*, the *sodA* gene (1.1-kb *EcoRI*-*KpnI* fragment; ref. 18). Labelling of probes and processing of blots was as described¹⁵. Lane 1, H37Rv; lane 2, strain 12, IC_{min} 1.6 µg ml⁻¹ INH; lane 3, B1453, IC_{min} > 50 µg ml⁻¹ INH²⁰; lane 4, strain 24, IC_{min} > 50 µg ml⁻¹ INH; lane 5, 79112, INH-sensitive²¹; lane 6, 12646, INH-sensitive²¹; lane 7, 79665, INH-sensitive²¹. INH susceptibilities were confirmed by inoculation of Lowenstein-Jensen slopes containing differing concentrations of INH.



multiple-drug-resistant strains in which there is a correlation between INH resistance and decreased catalase activity are particularly important because, owing to the contagiousness of tuberculosis, these strains pose a public health threat to both

HIV-infected and healthy individuals³. An improved understanding of the mechanisms of drug resistance will enable rapid tests for drug-resistance isolates to be developed and should facilitate the design of antituberculosis drugs. □

Received 13 May; accepted 2 July 1992.

- Murray, C. J. L., Styblo, K. & Rouillon, A. *Bull. Int. Union against Tuberculosis and Lung Disorders* **65**, 6-24 (1990).
- Perriens, J. H. *et al.* *Am. Rev. respir. Dis.* **144**, 750-755 (1991).
- Snider, D. E. & Roper, W. L. *New Engl. J. Med.* **326**, 703-705 (1992).
- Cohn, M. L., Kovitz, C., Oda, U. & Middlebrook, G. *Am. Rev. Tuberc.* **70**, 641-664 (1954).
- Middlebrook, G., Cohn, M. L. & Scheffer, W. B. *Am. Rev. Tuberc.* **70**, 852-872 (1954).
- Youatt, J. *Am. Rev. respir. Dis.* **99**, 729-749 (1969).
- Winder, F. G. in *The Biology of the Mycobacteria* Vol. 1 (eds Ratledge, C. & Stanford, J.) 353-438 (Academic, London, 1982).
- Snapper, S. B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 6987-6991 (1988).
- Snapper, S. B., Melton, R. E., Mustafa, S., Kieser, T. & Jacobs, W. R. *Molec. Microbiol.* **4**, 1911-1919 (1990).
- Heym, B. & Cole, S. T. *Res. Microbiol.* (in the press).
- Gayathri Devi, B., Shaila, M. S., Ramakrishnan, T. & Gopinathan, K. P. *Biochem. J.* **149**, 187-197 (1975).

- Triggs-Raine, B. L., Doble, B. W., Mulvey, M. R., Sorby, P. A. & Loewen, P. C. *J. Bact.* **170**, 4415-4419 (1988).
- Loprasert, S., Negoro, S. & Okada, H. *J. Bact.* **171**, 4871-4875 (1989).
- Mulvey, M. R., Sorby, P. A., Triggs-Raine, B. L. & Loewen, P. C. *Gene* **73**, 337-345 (1988).
- Zhang, Y. *et al.* *Infect. Immun.* **60**, 2160-2165 (1992).
- Quemard, A., Lacave, C. & Laneelle, G. *Antimicrob. Ag. Chemother.* **35**, 1035-1039 (1991).
- Shoeb, H. A., Bowman, B. U., Ottolenghi, A. O. & Merola, A. J. *Antimicrob. Ag. Chemother.* **27**, 404-407 (1985).
- Zhang, Y., Lathigra, R., Garbe, T., Catty, D. & Young, D. *Molec. Microbiol.* **5**, 381-391 (1991).
- Wayne, L. G. & Diaz, G. A. *Analyt. Biochem.* **157**, 89-92 (1986).
- Jackett, P. S., Aber, V. & Lowrie, D. J. *gen. Microbiol.* **104**, 37-45 (1978).
- Mitchison, D. A., Selkon, J. B. & Lloyd, J. J. *J. Path. Bact.* **86**, 377-386 (1963).

ACKNOWLEDGEMENTS. We thank W. R. Jacobs and P. C. Loewen for strains, and J. Grosset, R. Lathigra, T. Garbe and J. Ivanyi for discussion and encouragement. This work was supported by the World Health Organization Programme for Vaccine Development (D.Y. and S.C.), the Fondation pour la Recherche Médicale (S.C.), and the Association Française Raoul Follereau (S.C.).

Anatomy of a transcription factor important for the Start of the cell cycle in *Saccharomyces cerevisiae*

Michael Primig, Shanthini Sockanathan, Herbert Auer & Kim Nasmyth

Institute of Molecular Pathology, Dr Bohr-Gasse 7, A-1030 Vienna, Austria

ENTRY of yeast cells into the mitotic cell cycle (Start) involves a form of the CDC28 kinase that associates with G1-specific cyclins encoded by *CLN1* and *CLN2* (ref. 1). The onset of Start may be triggered by the activation of *CLN1* and *CLN2* transcription in late G1 (ref. 2). SWI4 and SWI6 are components of a factor (SBF) that binds the CACGAAA (SCB) promoter elements³⁻⁵ responsible for activation in late G1 of the *HO* endonuclease, *CLN1* and *CLN2* genes^{6,7}. A related factor (MBF) containing SWI6 and a 120K protein⁸ binds to the ACGCGTNA (MCB) promoter elements responsible for late G1-specific transcription of DNA replication genes⁹⁻¹². Nothing is known about how these heteromeric proteins bind DNA. We show here that SWI4 contains a novel DNA-binding domain at its N terminus that alone binds specifically to SCBs and a C-terminal domain that binds to SWI6. SWI4's DNA-binding domain is similar to an N-terminal domain of the *cdc10* protein that is a component of an MBF-like factor from *Schizosaccharomyces pombe*¹³ and is required for Start^{14,15}. An involvement of this kind of DNA-binding domain in transcriptional controls at Start may therefore be a conserved feature of eukaryotic cells.

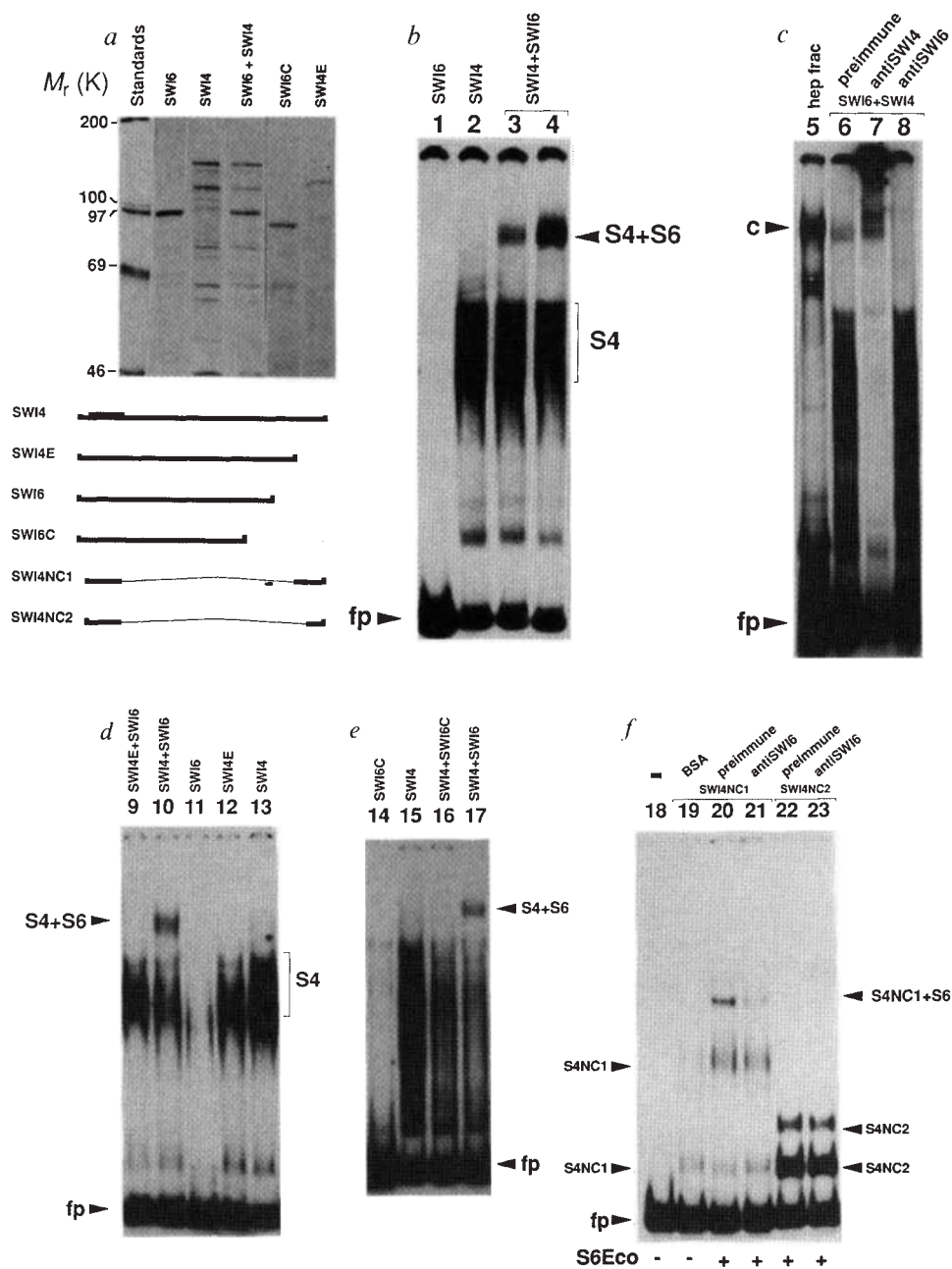
To determine whether SWI4 or SWI6 alone can bind SCBs, we translated both proteins in reticulocyte lysates. Full-length SWI6 (846 residues¹⁶) is made efficiently but much of SWI4 (1,094 residues^{4,17}) is either degraded or prematurely terminated (Fig. 1a). We tested the ability of the proteins to bind an oligonucleotide from the *CLN2* promoter (pCL2) that contains three potential SCBs and forms a complex with SBF isolated from yeast⁶. Using a gel retardation assay, we observe a heterogeneous set of SWI4:pCL2 complexes (Fig. 1b), all of which are recognized by a SWI4-specific antibody but not by preimmune serum (data not shown and Fig. 1c). The heterogeneity may be due to the variable size of the SWI4 protein. No complexes were observed using the SWI6 protein (Fig. 1b). That SWI4 but not SWI6 can bind SCB DNA is consistent with experiments showing that SWI4 overproduction allows *HO* to be transcribed without SWI6^{17,18} (but not vice versa) and that *CLN2* can be partially activated by SWI4 in *swi6* mutants^{6,19}.

Cotranslated SWI4 and SWI6 proteins form a new complex (with pCL2) containing both proteins that migrates with a mobility similar to that of the complexes formed by partially purified SBF from yeast (Fig. 1b, c). The complex formed with *in vitro* translated proteins seems to migrate slightly faster than that formed by yeast proteins and could conceivably lack a third component or modification. A truncated version of SWI4 lacking 144 amino acids from its C-terminal end (SWI4E) cannot form complexes with SWI6 although it still binds pCL2 (Fig. 1d). *HO* expression due to modest overproduction of such a protein in yeast is largely SWI6-independent¹⁷. Likewise, a version of SWI6 lacking its most C-terminal 89 amino acids (SWI6C) cannot form complexes on pCL2 with SWI4 (Fig. 1e). SWI4 and SWI6 might therefore interact by their C termini. This part of SWI6 is highly conserved in *Kluyveromyces fragilis*

FIG. 1 Binary and ternary complexes formed by SWI4 and SWI6 on SCB DNA.

fp, Free probe. **a**, Translation products labelled with 35 S-methionine and examined on SDS polyacrylamide gels²⁴. SWI4E refers to a truncated form of SWI4 lacking its C-terminal 144 amino acids due to digestion of the DNA template with *EcoRI*. SWI6C refers to a version of SWI6 that lacks its C-terminal 89 amino acids due to digestion of the DNA template with *ClaI*. SWI4's predicted M_r is 123K but it runs as a 150K protein. **b**, SWI4 binds DNA and recruits SWI6. Labelled DNA (0.5 ng) was incubated with 10 μ l reticulocyte lysates containing SWI6 (lane 1), SWI4 (lane 2) and both proteins (lanes 3 and 4). Four times less SWI4 RNA was added to the translation mix in lane 4 compared to lane 3. Note that the ternary complexes containing SWI4 and SWI6 are much more uniform than the binary complexes containing SWI4 alone. We presume that this is due to the involvement of SWI4's C-terminus in its interaction with SWI6 (see **d**); that is very few of the incomplete SWI4 molecules will be able to form ternary complexes. **c**, *In vitro* SWI4-SWI6 complexes are recognized by SWI4- and SWI6-specific antibodies and have a similar electrophoretic mobility to *in vivo* complexes (marked **c**). Yeast SBF (1 μ l) purified by heparin agarose chromatography (hep frac) was incubated with 0.5 ng labelled probe (lane 5). A reticulocyte lysate containing both SWI4 and SWI6 (lanes 6 to 8) was mixed with 1 μ l preimmune serum (lane 6), anti-SWI4 serum⁸ (lane 7) and anti-SWI6 serum⁵ (lane 8). **d**, The C-terminus of SWI4 is required for interaction with SWI6. Labelled probe (0.5 ng) was incubated with 10 μ l lysates containing SWI4 (lane 13), SWI4E (lane 12), SWI6 (lane 11) as well as SWI4 and SWI4E cotranslated with SWI6 (lanes 10 and 9). Lysate alone (not shown) gave the same pattern as SWI6 protein. **e**, The C terminus of SWI6 is required for interaction with SWI4. Labelled probe (0.5 ng) was incubated with 10 μ l lysates containing SWI6C (lane 14), SWI4 (lane 15), SWI6C plus SWI4 (lane 16), and SWI6 plus SWI4 (lane 17). **f**, The C terminus of SWI4 is sufficient for interaction with SWI6. Labelled probe (0.5 ng) was incubated with a control lysate (lane 18), with SWI4NC1 (lanes 19–21) and SWI4NC2 (lanes 22 and 23). SWI6 (0.2 μ g) purified from *E. coli*⁵ (S6Eco) was added in lanes 20–23 as indicated at the bottom of the panel. Preimmune serum (1 μ l) diluted 1:10 was added in lanes 20 and 22. SWI6 antiserum diluted 1:10 was added in lanes 21 and 23. The SWI4NC-dependent complexes are marked S4NC1 and S4NC2. The SWI4NC1:SWI6 complexes are only partially abolished by SWI6 anti-serum because of the large amount of SWI6 protein present.

METHODS. RNA synthesized by T7 RNA polymerase²⁵ was added to a rabbit reticulocyte lysate system (Promega L4210) at concentrations of 10–50 μ g ml⁻¹. BSA, 0.5 μ g ml⁻¹ salmon sperm DNA at a 30-fold molar excess of 32 P-GpppG CAP analogue (Pharmacia) versus rGTP. Labelled oligonucleotide (0.5 ng) and 3–15 μ l reticulocyte lysate was incubated in 20 mM Tris pH 7.5, 3 mM MgCl₂, 1 mM DTT, 50 mM NaCl, 0.1 mM EDTA, 5 mM spermidine, 50 μ g ml⁻¹ BSA, 0.5 μ g ml⁻¹ salmon sperm DNA and 5% glycerol for 15–30 min in a final volume of 20 μ l. For each μ l of lysate, 0.5 μ g Poly d(I), d(C) from Pharmacia was added. Complexes were analysed on 4% 20:1 acrylamide:bisacrylamide gels in 0.25 $\times$ TBE buffer (1 $\times$ TBE is 89 mM Tris, 89 mM borate and 2.4 mM EDTA) that were prerun at 10 V cm⁻¹ for 1 h. In competition experiments, unlabelled oligonucleotides were added at a 500-

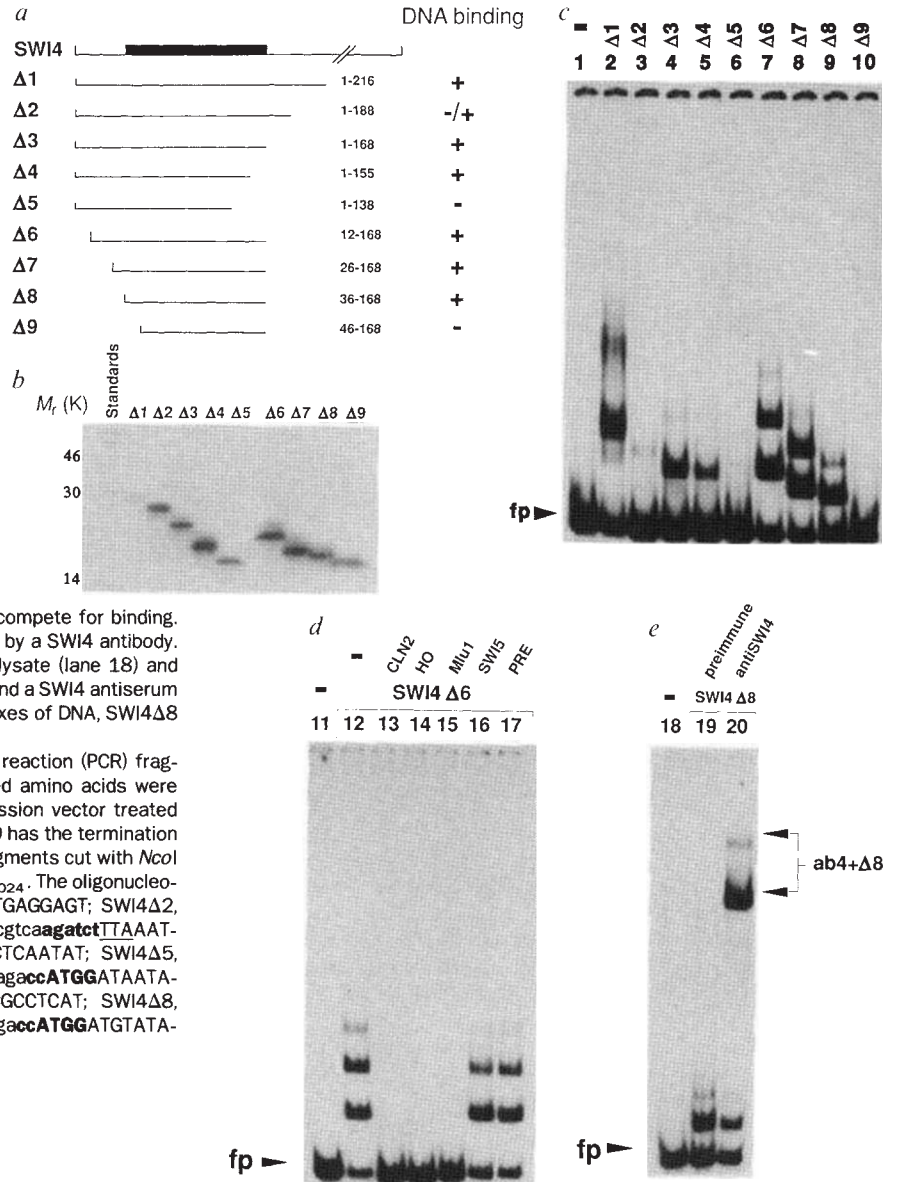


fold molar excess before addition of the lysate samples. Antibodies were added last (10 min before electrophoresis). Gels were run at 10 V cm⁻¹ dried and exposed overnight at -80 °C with an intensifying screen. The following oligonucleotides were used to construct the *in vitro* expression vectors of SWI4, SWI6, SWI4NC1 and SWI4NC2: SWI6-1, taattcgaattc-ccATGGCGTTGGAAGAAG; SWI6-2, aaaatttcgaaggaggaaatggCCAGTCTC; SWI4-1, gatcggggCCATTTGATGTTTTGATATCAA; SWI4-2, tcgatcagatct-TTATGCGTTTGCCCTCAA; SWI4-A, ttggcgtgcaGTGTAGCTCGATGGAGAAG; SWI4-B, gatcgactgcaAAAAAGCAAAATTTTATTC; SWI4-C, gatcgactgcaG-AGACGTTGAGACTAGCAA; To construct the SWI6 expression plasmid, pT7₇₅₅₋₁₀₂₄ was linearized with *NcoI* and *BamHI* and triple ligated to a SWI6 PCR fragment cut with *NcoI* and *MscI* plus a *MscI*-SWI6-*BglII* restriction fragment from pBd159 (a SWI6 *HindIII* fragment cloned into the *HindIII* site of pLC19H). To obtain the SWI4 vector, pT7₇₅₅₋₁₀₂₄ was digested with *NcoI*, end repaired with Klenow and redigested with *BamHI*. The linearized plasmid was ligated to a SWI4 PCR fragment that was digested with *HaeIII* and *BglII*. The SWI4NC1 fusion was a triple ligation of 2 PCR fragments obtained with the oligonucleotides SWI4Δ8 and SWI4-A as well as SWI4-B and SWI4-2; the former was cut with *NcoI* and *PstI*, the latter with *PstI* and *BglII* and ligated to pT7₇₅₅₋₁₀₂₄ treated as described for SWI6. SWI4NC2 was produced the same way except that the oligonucleotide SWI4-C was used instead of SWI4-B.

FIG. 2 Defining SWI4's DNA-binding domain. *a*, Deleted versions of SWI4 and a summary of their DNA binding. The black bar indicates the DNA-binding domain. Amino-acid end points are indicated by numbers on the right. *b*, Translation products of the above deletions. Size markers (K) on the left. The low abundance of $\Delta 1$ might be a loading artefact. *c*, DNA-binding assays. Labelled probe (0.5 ng) was incubated with 4 μ l reticulocyte lysates containing the translation products of RNAs encoding viral coat proteins (lane 1) and SWI4 mutants ($\Delta 1$ – $\Delta 9$, lanes 2–10). fp, Free probe.

d, SWI4 DNA binding is competed by oligonucleotides containing SCBs. Labelled probe (0.5 ng; pCL2) was mixed with a control lysate (lane 11) and SWI4 $\Delta 6$ (lanes 12–17). Unlabelled competitors were added at a 500-fold molar excess in lanes 13–17. The oligonucleotide competitors are: unlabelled probe, pCL2⁵ (lane 13); an *HO* promoter fragment containing three SCBs, sRS2⁵ (lane 14); a *TMP1* promoter fragment containing two MCBs, MCB-TMP1⁸ (lane 15); an *HO* promoter fragment containing a SWI5-binding site²⁶ (lane 16); and a fragment containing PREs bound by STE12 (lane 17) (M.P. and G. Ammerer, manuscript in preparation). Note that MCBs as well as SCBs can compete for binding. *e*, SWI4 $\Delta 8$:pCL2 gel retardation complexes are shifted by a SWI4 antibody. Labelled probe (0.5 ng) was incubated with a control lysate (lane 18) and SWI4 $\Delta 8$ (lanes 19 and 20). A preimmune serum (1 μ l) and a SWI4 antiserum (1 μ l)⁸ were added in lanes 19 and 20. Ternary complexes of DNA, SWI4 $\Delta 8$ protein and SWI4 antibody are marked (ab4 + $\Delta 8$).

METHODS. As described for Fig. 1. Polymerase chain reaction (PCR) fragments with either ATG or TAA codons at the indicated amino acids were digested with *Hae*III and *Bgl*II and ligated to an expression vector treated as described for SWI4 to yield SWI4 $\Delta 1$ – $\Delta 5$; SWI4 $\Delta 6$ – $\Delta 9$ has the termination codon from SWI4 $\Delta 3$ and were made by cloning PCR fragments cut with *Nco*I and *Bgl*II between the *Nco*I and *Bam*HI sites of PT7_{755–1024}. The oligonucleotides used were: SWI4 $\Delta 1$, tcgatcagatctTTAGAACTGTGAGGAGT; SWI4 $\Delta 2$, tcgatcagatctTTAATTCGAGCTGTAGT; SWI4 $\Delta 3$, tcgtcaagatctTTAAATGTAGCTCGATGG; SWI4 $\Delta 4$, tcgtcaagatctTTAAGTCTCCTCAATAT; SWI4 $\Delta 5$, tcgatcagatctTTAATCGAATTGAAAAGT; SWI4 $\Delta 6$, actagaccATGGATAATCAATACACAG; SWI4 $\Delta 7$, actagaccATGGTTCTTTTGGCGCCTCAT; SWI4 $\Delta 8$, actagaccATGGTGATTGAAATAGCTACG; SWI4 $\Delta 9$, actagaccATGGATGTATACGAATGCTATA.



(M. Neuberg, personal communication) and is similar to an equivalent domain of the *cdc10* protein (Cdc10)¹⁶.

To define SWI4's DNA-binding domain more precisely, we compared the binding activity of various truncated forms (Fig. 2*a*). SWI4 polypeptides lacking all amino acids C-terminal to position 155 can still form binary complexes, as can polypeptides lacking the N-terminal 36 amino acids (Fig. 2*a, b, c*). But deleting C-terminal to residue 36 or N-terminal to residue 155 abolishes binding. A SWI4 polypeptide containing only residues 36 to 168 ($\Delta 8$) forms complexes almost as efficiently as larger forms. The only anomaly of this deletion series is $\Delta 2$ which binds poorly. The complexes formed by $\Delta 8$ contain SWI4 protein (Fig. 2*e*) and those formed by $\Delta 6$ are competed by oligonucleotides containing SCBs from the *CLN2* and *HO* promoters but not by unrelated oligonucleotides (Fig. 2*d*). An oligonucleotide containing MCBs from the *TMP1* promoter also competes. The amino acids 36–155 that form SWI4's DNA-binding domain are, in contrast to much of SWI4's primary sequence, highly conserved in a related protein for *K. lactis* (M. Neuberg, personal communication) and are sufficient for binding to the *HO* promoter *in vivo* (N. Jones, personal communication). But this domain alone cannot activate *HO* transcription¹⁷. Most of the SWI4 polypeptides capable of binding to pCL2 give rise to two

and often three complexes. The number of complexes is related to the number of SCBs in the DNA (Fig. 3*a, c*). An oligonucleotide containing only a single SCB (pCL2 $\Delta 2$) forms only one type of complex with either SWI4 $\Delta 7$ or SWI4 $\Delta 8$, whereas an oligonucleotide containing two SCBs forms two types of complex (Fig. 3*a*). Cotranslation of SWI4 $\Delta 1$ and SWI4 $\Delta 8$ produced complexes with the expected mobilities but did not produce any complexes with an intermediate mobility indicative of hybrids (data not shown). These data imply that SWI4 binds as a monomer to single SCBs and that it can bind simultaneously to all three SCBs in pCL2. To determine the binding site of SWI4 on pCL2 $\Delta 2$ more precisely, we have determined at which adenine or guanine residues carbethoxylation interferes with binding. Interference is restricted to bases in the original SCB consensus sequence²⁰ (Fig. 3*b*). The pattern suggests that SWI4 makes contacts along an extensive stretch of the major groove.

To test whether SWI4's C terminus is sufficient for forming ternary complexes with SWI6 on DNA, we analysed complex formation with deleted versions of SWI4 lacking the central region due to fusion of SWI4's DNA-binding domain directly to C-terminal sequences. Neither the DNA-binding domain alone (data not shown) nor a fusion containing the C-terminal 65 residues (SWI4NC2) can recruit SWI6 but a larger fusion

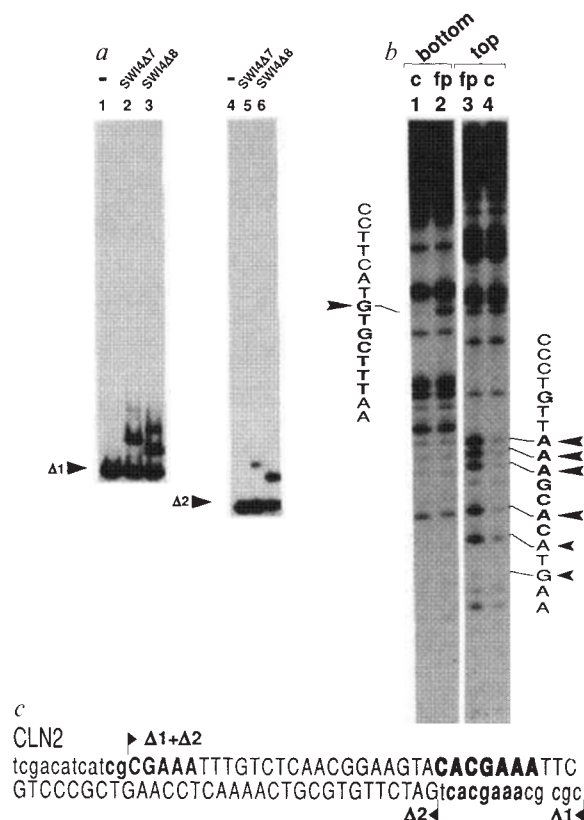


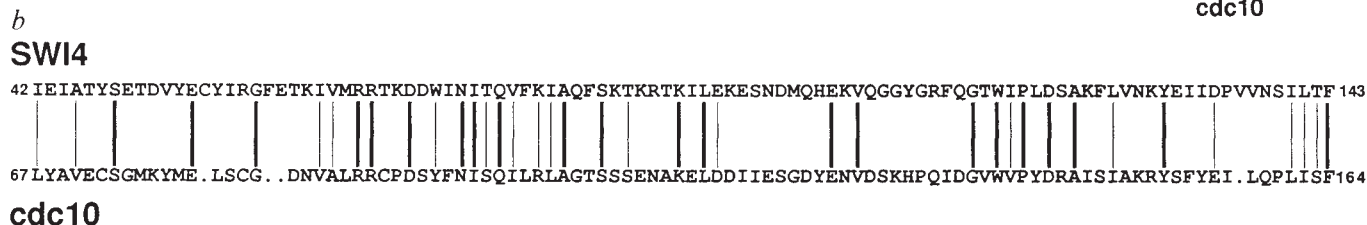
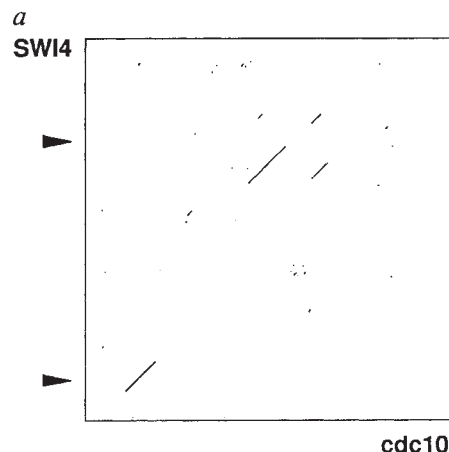
FIG. 3 The SWI4's DNA-binding domain recognizes a single SCB element. **a**, End-labelled pCL2Δ1 (0.5 ng) (lanes 1–3) and pCL2Δ2 (lanes 4–6) probes were incubated with control lysates (lanes 1 and 4) as well as SWI4Δ7 (lanes 2 and 5) and SWI4Δ8 (lanes 3 and 6). **b**, Carboxymethylated DEPC-treated oligonucleotide (pCL2Δ2) containing a single SCB from the *CLN2* promoter was incubated with SWI4Δ8. Unbound (fp, lanes 2 and 3) and complexed DNAs (c, lanes 1 and 4) were separated by gel electrophoresis as in **a**, and the hydrolysed products analysed on a 10% sequencing gel⁸. Interfering bases are indicated by arrowheads. **c**, The sequences of pCL2, pCL2Δ1 and Δ2. The pCL2 sequence is shown on two lines and the end points of Δ1 and Δ2 are flagged. Large bases are those in pCL2Δ2 whose SCB is in bold letters. Small bases are deleted at the 5'-end of pCL2Δ1 or both at the 5'- and 3'-end of pCL2Δ2.

containing 146 C-terminal residues (SWI4NC1) allows SWI6 to form complexes on SCB DNA (Fig. 1f). Our data imply that SWI4's terminal 149 amino acids are both necessary and sufficient (with the DNA-binding domain) for recruiting SWI6 into ternary complexes *in vitro* but do not exclude other sequences also contributing.

We have found only a single protein that exhibits significant similarity to SWI4's DNA-binding domain. This is the Cdc10 protein from *Schizosaccharomyces pombe*²¹. A DIAGON comparison of SWI4 and Cdc10 (Fig. 4a) reveals two regions of similarity, one in the N-terminus that has not previously been

recognized and a second in the middle of both proteins that corresponds to the two 33 amino acid 'Notch' repeats also found in SWI6¹⁶ and many other proteins^{22,23}. The N-terminal region of similarity extends from residues 42 to 143 in SWI4, which corresponds almost exactly to its DNA-binding domain. Recently, Cdc10 has been shown to be part of a transcription factor¹³ that binds to the Start-dependent MCB element found in the promoters of many DNA replication genes in *S. cerevisiae*^{9–12}. We suggest that this region of Cdc10 may be involved in the recognition of MCBs. It seems likely that the *S. cerevisiae* p120 protein that forms complexes with SWI6 on

FIG. 4 The region of SWI4 covering its DNA-binding domain is conserved in Cdc10 from *S. pombe*. The results of a DIAGON (a) and a peptide-sequence alignment (b) obtained with the GCG package²⁷. The sequences of SWI4 and Cdc10 were aligned with the BESTFIT program. The DOTPLOT parameters of the COMPARE program were window 35 with a stringency of 18.0. Identical residues and conservative changes are indicated by thick and thin lines, respectively. A SWI4 gene cloned and sequenced by us¹⁷ had an identical sequence to that published by Andrews and Herskowitz⁴. One problem with the proposal that the N-terminal domain of Cdc10 may bind DNA is that a fragment of Cdc10 lacking this domain is capable at high copy number of suppressing the lethality of a temperature-sensitive Cdc10 mutant²¹. It is possible that the mutation whose phenotype is suppressed may lie in the C-terminal half of Cdc10 and the suppressing fragment may only enable the mutant protein with an intact DNA-binding domain to function.



MCB elements⁸ will also turn out to have a similar DNA-binding domain. SCBs and MCBs are related in sequence and when present at high concentration can compete with each other for the binding of SWI4/SWI6 (SBF) and p120/SWI6 (MBF) complexes (ref. 8, Fig. 2d, and data not shown). Nevertheless, SBF and MBF bind preferentially to SCB and MCB elements, respectively^{6,8}.

Apart from sharing two Notch repeats, the C-terminal half of Cdc10 is unlike that of SWI4. It instead has extensive similarity to the C-terminal half of SWI6¹⁶. The Cdc10 protein therefore has the appearance of a hybrid between SWI4 and SWI6. In

contrast to Cdc10, the N terminus of SWI6 has very little resemblance to SWI4. It may therefore not possess a functional DNA-binding domain, which is consistent with it not binding DNA in *in vitro* assays (Fig. 1b). One explanation for the hybrid character of Cdc10 is that all of these proteins are descended from a Cdc10-like ancestor, one of whose descendants (SWI4) has, apart from the Notch repeats, diverged extensively in its C terminus and another (SWI6) has diverged extensively in its N terminus. The conservation of SWI4's DNA-binding domain suggests that it may be found in Start-dependent transcription factors in many eukaryotic organisms. □

Received 8 April; accepted 18 June 1992.

- Richardson, H. E., Wittenberg, C., Cross, F. & Reed, S. I. *Cell* **59**, 1127–1133 (1989).
- Wittenberg, C., Sugimoto, K. & Reed, S. I. *Cell* **62**, 225–237 (1990).
- Breeden, L. & Nasmyth, K. *Cell* **48**, 389–397 (1987).
- Andrews, B. J. & Herskowitz, I. *Nature* **342**, 803–833 (1989).
- Taba, M. R., Muroff, I., Lydall, D., Tebb, G. & Nasmyth, K. *Genes Dev.* **5**, 2000–2013 (1991).
- Nasmyth, K. & Dirick, L. *Cell* **66**, 995–1013 (1991).
- Ogas, J., Andrews, B. J. & Herskowitz, I. *Cell* **66**, 1015–1026 (1991).
- Dirick, L., Moll, T., Auer, H. & Nasmyth, K. *Nature* **357**, 508–513 (1992).
- Pizzagalli, A., Valsasini, P., Plevani, P. & Lucchini, G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3772–3776 (1988).
- McIntosh, E. M., Atkinson, T., Storms, R. K. & Smith, M. *Molec. cell. Biol.* **11**, 329–337 (1991).
- Gordon, C. B. & Campbell, J. L. *Proc. natn. Acad. Sci. U.S.A.* **68**, 6058–6062 (1991).
- Lowndes, N. F., Johnson, A. L. & Johnson, L. M. *Nature* **350**, 247–250 (1991).
- Lowndes, N. F., McInerney, C. J., Johnson, A. L., Fantes, P. A. & Johnston, L. H. *Nature* **355**, 449–453 (1992).
- Nurse, P., Thuriaux, P. & Nasmyth, K. *Molec. Gen. Genet.* **146**, 167–178 (1976).
- Nurse, P. & Bisset, Y. *Nature* **292**, 558–560 (1981).
- Breeden, L. & Nasmyth, K. *Nature* **329**, 651–654 (1987).

- Sockanathan, S. thesis, Univ. Cambridge (1991).
- Breeden, L. & Mikesell, G. E. *Genes Dev.* **5**, 1183–1190 (1991).
- Moll, T., Dirick, L., Auer, H. & Nasmyth, K. *J. Cell Science* (in the press).
- Nasmyth, K. *Cell* **42**, 213–223 (1985).
- Aves, S. J., Durkacz, B. W., Carr, A. & Nurse, P. *EMBO J.* **4**, 457–463 (1985).
- Lux, S. E., John, K. M. & Bennett, V. *Nature* **344**, 36–42 (1990).
- Thompson, C. C., Brown, T. A. & McKnight, S. L. *Science* **253**, 726–768 (1991).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Van der Werf, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2330–2334 (1986).
- Stillman, D. J., Bankier, A. T., Seddon, A., Groenhout, E. G. & Nasmyth, K. *EMBO J.* **7**, 485–494 (1988).
- Devereux, J., Haeblerli, P. & Smithies, O. *NAR* **12**, 387–395 (1984).
- Sturm, R., Baumruker, T., Franua, B. R. & Herr, W. *Genes Dev.* **1**, 1147–1160 (1987).

ACKNOWLEDGEMENTS. We thank H. Tkadletz for help in preparing the figures, R. Kurzbaier and G. Schaffner and co-workers for oligonucleotide synthesis and DNA sequencing, M. Nicklin for providing T7 RNA polymerase, T. Moll for antibodies against SWI4, C. Koch, B. Errede and L. Ballou for comments on the manuscript, G. Ammerer in whose laboratory this work was done and encouraged and R. Treisman for plasmid PT7⁷⁵⁵⁻¹⁰²⁴. M.P. was supported by an FWF grant to G.A. This work was supported by the Austrian Industrial Research Promotion Fund.

Candidate proto-oncogene *bcl-3* encodes a subunit-specific inhibitor of transcription factor NF- κ B

F. Gregory Wulczyn, Michael Naumann & Claus Scheidereit*

Otto Warburg Laboratorium, Max Planck Institut für Molekulare Genetik, Ihnestr. 73, 1000 Berlin 33, Germany

THE NF- κ B subunits p50 and p65 and the product of the *rel* proto-oncogene are members of a growing class of transcription factors with a unique DNA-binding and dimerization domain^{1–13}. Nuclear transfer of each of these factors is controlled by cytoplasmic inhibitors, and regulated by specific stimuli. The inhibitors I κ B- α and - β and pp40 recognize either p65 or the c-*rel* protein^{14–16}. We show here that the proto-oncogene *bcl-3*, believed to be involved in certain human B-cell leukaemias¹⁷, encodes a protein that functions as an I κ B-like molecule for native NF- κ B but is specific for the p50 subunit. The ankyrin repeat domain of the *bcl-3* product is shown to mediate complex formation with NF- κ B dimers by contacting the conserved dimerization domain of NF- κ B.

On the basis of structural similarities in a repeat domain of the *bcl-3* protein (Bcl-3) and a region in the C-terminal domain of p105 (refs 3–6) and MAD-3/I κ B- α (ref. 18), we predicted as a common function the interaction with Rel-like transcription factors and reported that DNA binding of p50 is inhibited by the repeat domain of p105 and by Bcl-3 (ref. 19). We now report that native NF- κ B is a target for Bcl-3.

In a band-shift assay with nuclear extracts from HeLa cells, Bcl-3 specifically inhibited nuclear NF- κ B induced with the phorbol ester phorbol myristate acetate (PMA) (Fig. 1a, lanes 3 and 4), but did not affect endogenous octamer-binding transcription factor 1 (OTF-1) binding (lanes 5, 6). The inhibition

was achieved using an amount of Bcl-3 comparable to the amount of MAD-3/I κ B (ref. 18) necessary for inhibition of native NF- κ B (not shown), demonstrating that Bcl-3 is an alternative form of I κ B. We next analysed the subunit specificity of Bcl-3 for various Rel-like factors. DNA binding of p50, or of truncated p50 containing only the conserved Rel domain (p50 Δ), was effectively inhibited by Bcl-3 (Fig. 1b, lanes 1 to 4). Both the c-*rel* protein and truncated p65 were recognized (lanes 5 to 8), but less efficiently. In contrast, the *Drosophila* morphogen *dorsal* was not a target (lanes 9, 10). The affinities of Bcl-3 and MAD-3/I κ B were compared for both subunits of NF- κ B (Fig. 1c). MAD-3 has a strong preference for p65, but can at high concentrations also inhibit p50 (lanes 5–8, 13–16). Bcl-3 clearly has an inverse specificity, with a higher affinity for p50 than for p65 (lanes 1–4, 9–12). These results establish that Bcl-3 is a previously undetected I κ B species with a novel specificity, and suggest that I κ B molecules recognize the Rel domain in a conserved manner.

A deletional analysis of the Bcl-3 molecule (Fig. 2) revealed that the ankyrin repeat domain is necessary and sufficient for inhibition (Fig. 2b, lanes 1 to 4), and that removal of either part of the first repeat (lanes 5 and 6), or of the seventh and part of the sixth (lanes 7 and 8), results in inactivation. This is in contrast to erythrocyte ankyrin and GA binding protein β in which a subset of the ankyrin repeats can mediate protein–protein interaction^{20,21}, but is similar to the ankyrin repeat domains of p105 and MAD-3 (ref. 19; and E. Hatada and C.S., unpublished observation).

To determine which domain of p50 is contacted by Bcl-3, deletion mutants of p50 were assayed for interaction with immobilized Bcl-3 (Fig. 3). At the C terminus, amino acids up to and including the conserved nuclear transfer signal could be deleted (p50 Δ and C1) without affecting binding to the Bcl-3 matrix, but further deletion of 18 (C2) or 38 (C3) amino acids abrogated interaction (Fig. 3b). For unknown reasons, N-terminal deletions led to low levels of nonspecific binding to the control matrix (N2 and N3). Nevertheless, deletion of either the N-terminal 75 (N1) or 201 (N2) amino acids did not reduce binding to Bcl-3 compared to full-length p50; and the level of specific binding was significantly higher than to the immobilized

* To whom correspondence should be addressed.

FIG. 1 *a*, Inhibition of DNA binding of nuclear NF- κ B by recombinant Bcl-3. Nuclear extracts of non-stimulated (lanes 1, 2, 5, 6) or PMA-induced (lanes 3, 4) HeLa cells were assayed with bacterially expressed Bcl-3 (lanes 2, 4, 6) or control protein (lanes 1, 3, 5) in gel retardation assays using either κ B (lanes 1–4)²⁴ or H2B octamer (lanes 5, 6)¹⁹ probes. The positions of NF- κ B, OTF-1, and a nonspecific complex (ns) are indicated.

b, Bcl-3 interaction with vertebrate Rel-family proteins. Gel retardation assays were run with 5 pmol p50 (amino acids 1 to 462); 3 pmol truncated p50 (p50 Δ , amino acids 1 to 368); 30 pmol c-rel protein; 0.6 pmol truncated p65 (p65 Δ , amino acids 1 to 312); or 0.5 pmol dorsal protein as indicated. Reactions contained either control protein (lanes 1, 3, 5, 7, 9) or 20 pmol bacterial bcl-3 protein (lanes 2, 4, 6, 8, 10). Probes were κ B site (lanes 1–8) and Zerknüllt site²⁵ (lanes 9, 10). *c*, Inverse specificities of MAD-3/I κ B α and Bcl-3 to p50 and p65. Binding reactions contained either 1 pmol bacterial p50 or 0.6 pmol p65 Δ , as indicated. Increasing amounts of bacterial Bcl-3 or MAD-3 were added: 0 pmol (lanes 1, 5, 9, 13), 0.6 pmol (lanes 2, 6, 10, 14), 3 pmol (lanes 3, 7, 11, 15), and 15 pmol (lanes 4, 8, 12, 16), as indicated. Only protein-DNA complexes are shown.

METHODS. HeLa cells were stimulated for 3 h with 40 nM PMA; nuclear extracts were prepared as described²⁶. Bacterial bcl-3 (clone B2, Fig. 2) MAD-3, p50, p50 Δ and murine c-rel¹⁵ were expressed and purified by preparative SDS-PAGE as described¹⁹. The expression construct for MAD-3 contained coding sequences from amino acid 52 to 317 cloned into the pET3c vector (Novagen). Control protein was prepared as for Bcl-3 from cells containing pET3c vector without insert¹⁹; p65 Δ and dorsal were prepared as described²⁵.

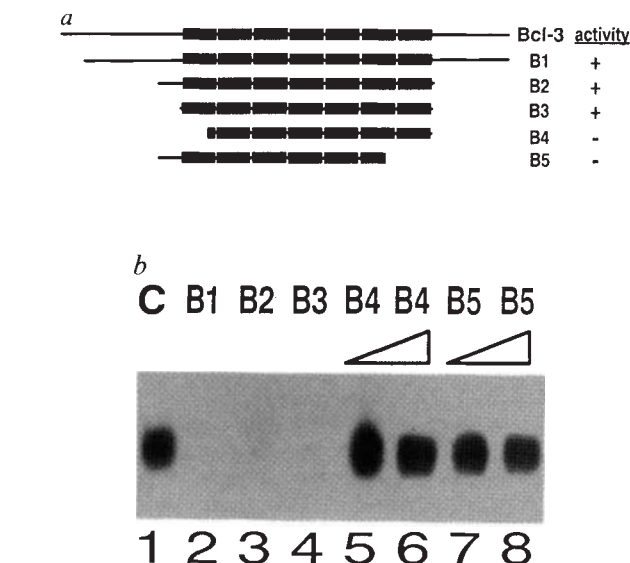
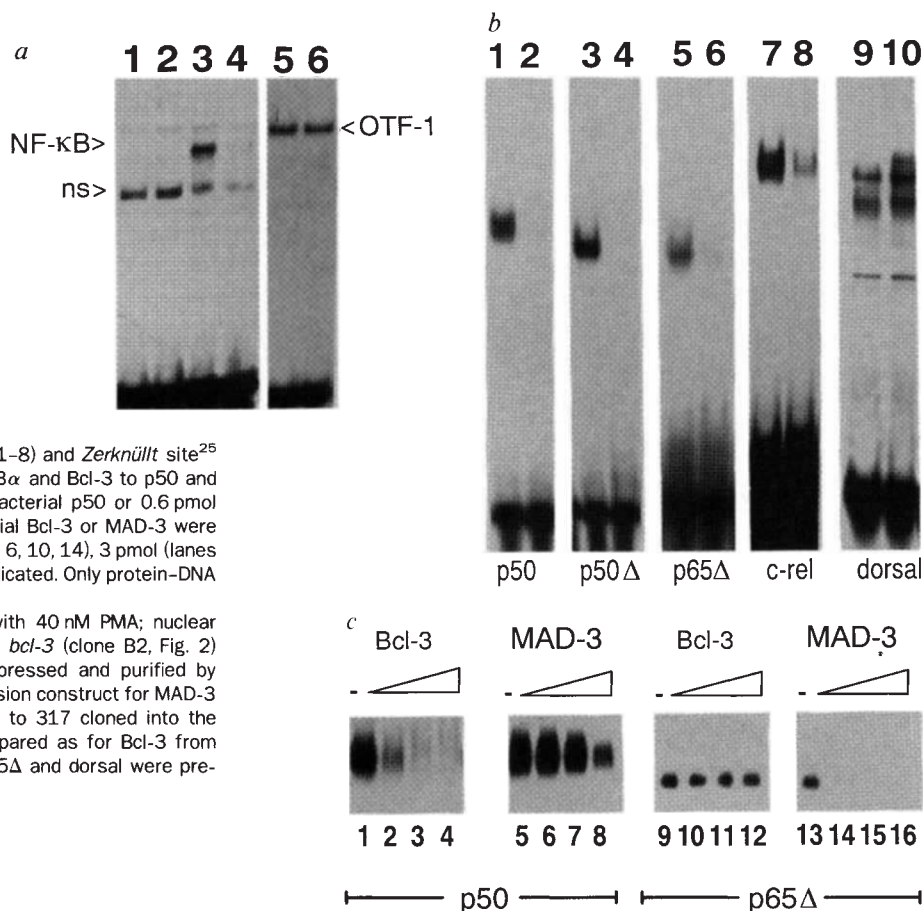


FIG. 2 Delineation of the inhibitory domain in Bcl-3. *a*, Deletion mutants of bcl-3 were cloned into pET3c, with the exception of B1, which was cloned into pRSETA (Invitrogen). The expressed peptides extend from amino acid 26 to 454 (B1); 94 to 383 (B2); 128 to 383 (B3); 155 to 383 (B4); 94 to 325 (B5) of Bcl-3 (ref. 17). Each ankyrin repeat is represented by a filled box. *b*, Gel retardation assay using bcl-3 mutants described in *a*. Reactions contained 2 pmol bacterial p50 together with control extract (lane 1); 8 pmol B1 (lane 2); 8 pmol B2 (lane 3); 8 pmol B3 (lane 4); 8 pmol (lane 5) or 40 pmol (lane 6) B4; 8 pmol (lane 7) or 40 pmol (lane 8) B5. Only protein-DNA complexes are shown. Mutant proteins were partially purified as described in²⁷; a control extract (C) was prepared from cells containing only pRSETA.

glutathione-S-transferase control (Fig. 3*b*). Further deletion to amino acid 283 (N3) reduced binding to Bcl-3 to the same extent as the control, indicating the loss of residues essential for the interaction.

These results and the DNA-binding activities of the deletion mutants are summarized in Fig. 3*a*. It is surprising that both the N-terminal domain of p50, required for DNA binding, and the nuclear transfer signal are dispensable for binding by Bcl-3. Instead, the interaction domain defined here is very similar to the region that is essential for p50 dimerization²². The relation between p50 dimerization and Bcl-3 binding was therefore analysed by crosslinking with glutaraldehyde.

Analysed separately, p50 is dimeric and Bcl-3 is a monomer (Fig. 4*a*, lanes 1, 2 and 5, 6); the Bcl-3 molecule used was encoded by construct B2 (Fig. 2*a*). In an equimolar mixture two new complexes are formed with relative molecular masses of ~115K and 150K (lanes 3 and 4). The sizes of the complexes correspond to either 1 or 2 Bcl-3 molecules bound to a p50 dimer, respectively. Heterodimeric complexes were not observed. To confirm the presence of Bcl-3 in both complexes, we used the full-length Bcl-3 construct (B1; Fig. 2*a*) to achieve crosslinking to p50, and probed the blot with an antiserum specific for Bcl-3. Again two complexes were seen and their M_r s were shifted to about 130K and 180K as expected (Fig. 4*b*, lanes 1 and 2). Thus, Bcl-3 efficiently binds to p50 dimers without dissociating them, and two Bcl-3 molecules can be accommodated per dimer. These results imply that the ability of I κ B molecules to inhibit DNA binding and nuclear transfer is presumably indirect, probably through steric or allosteric effects. Furthermore, in the cell it is feasible that heterodimeric Rel proteins are engaged by combinations of inhibitors.

Taken together, the deletion analysis and the crosslinking data demonstrate that Bcl-3 contacts the dimerization domain

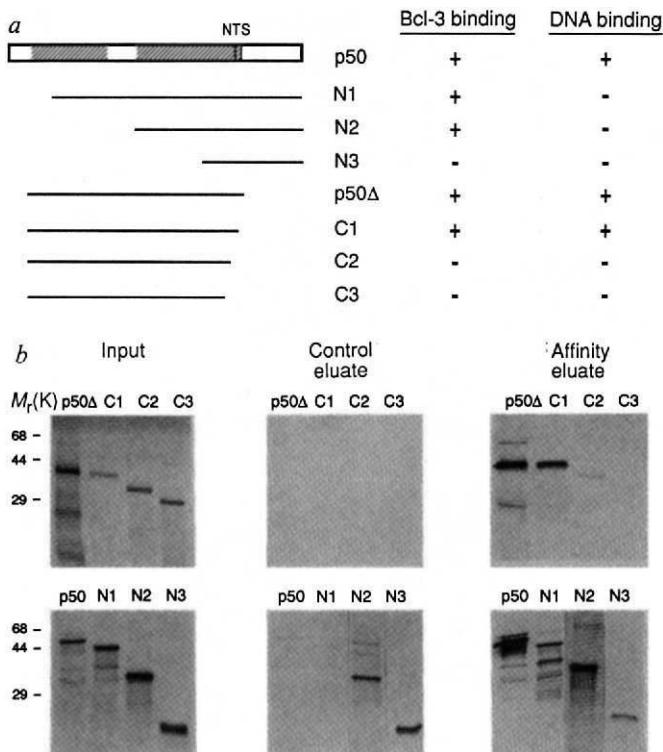


FIG. 3 Mapping of the interaction domain on p50 for Bcl-3 by affinity chromatography. **a**, Schematic representation of p50 mutants and summary of DNA-binding and Bcl-3-association activities. Hatching indicates homology with Rel. **b**, Autoradiograms of SDS-PAGE: Results obtained with C-terminal deletions are shown in the top panels, N-terminal deletions in the bottom panels. Left, relative protein amounts applied to the matrix; centre, protein eluted from control matrix; right, affinity matrix eluate. In brief, Bcl-3-glutathione-S-transferase fusion protein was coupled to glutathione-agarose²⁸. Substrates for *in vitro* transcription were generated by PCR, 5' primers contained signals for T7 RNA polymerase and translation initiation. The fragments generated encoded amino acids 37 to 462 (p50); 78 to 462 (N1); 202 to 462 (N2); 286 to 462 (N3); 37 to 368 (p50Δ); 37 to 362 (C1); 37 to 344 (C2); 37 to 328 (C3) of p50 (ref. 6). ³⁵S-labelled products of *in vitro* translation reactions (input) were preincubated with glutathione-S-transferase beads, and then adsorbed either to coupled Bcl-3-glutathione-S-transferase beads, or to coupled glutathione-S-transferase protein (control matrix). Conditions for binding and elution have been described²⁸.

of p50, and that the dimeric form is the target for Bcl-3. Alternatively, the requirement for dimerization could be structural, with only a subregion of the dimerization domain contacted directly. Nevertheless, we postulate that in general the Rel type dimerization domain is the target for the ankyrin repeat domains of IκB molecules. Subunit specificity must then be determined by variable residues within these domains, leading to subgroups of inhibitors with preferences either for p50 (refs 19, 23), and

perhaps other closely related Rel-like factors, or for p65 and c-rel protein^{14-16,18}. Inhibitors with different subunit specificities might allow differential induction of particular Rel homo- and heterodimers, or account for cell-type specific differences in the repertoire of inducible Rel-like factors. The observation that both upregulation of Bcl-3 in the 14; 19 translocation¹⁷, and loss of an ankyrin domain in Iy-10 (ref. 10) are associated with certain lymphomas suggest that a proper balance of inhibitors and activators is essential for normal cell function. □

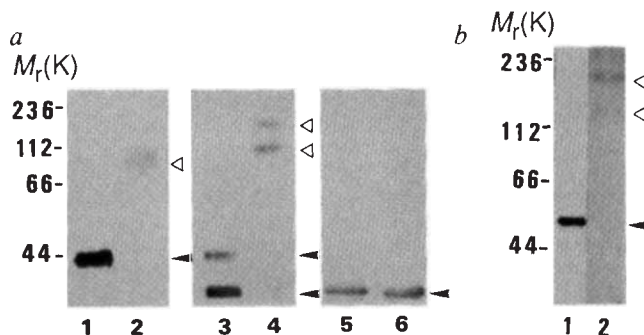


FIG. 4 Physical interaction study of p50 and Bcl-3 by glutaraldehyde cross-linking. **a**, p50Δ (42K) or bcl-3 construct B2 (35K) (lanes 1, 2 and 5, 6) or a mixture of both (lanes 3, 4) were incubated with (lanes 2, 4 and 6) or without glutaraldehyde (lanes 1, 3 and 5) and analysed in a western blot. M_r markers are indicated. Filled arrows indicate protein monomers; open arrows homo- or heteromeric protein complexes. **b**, p50Δ (42K) and bcl-3 construct B1 (48K) were incubated without (lane 1) or with glutaraldehyde (lane 2) and analysed as above using a Bcl-3-specific antibody. **METHODS.** The p50Δ (amino acids 1-368) and bcl-3 construct B1 (amino acids 26-454) and B2 (amino acids 94-383) proteins were gel-purified and renatured¹⁹. For crosslinking, 1 pmol each was incubated with 0.0025% glutaraldehyde for 1 h at 22 °C. Protein samples were separated by SDS-PAGE²⁹ and transferred to a PVDF membrane (Millipore). Protein complexes were visualized either by an antibody against an N-terminal vector-derived tag (Novagen) (**a**), or an affinity-purified serum raised against Bcl-3 (**b**).

Received 8 April; accepted 24 June 1992.

- Wilhelmsen, K. E., Eggleston, K. & Temin, H. M. *J. Virol.* **52**, 172-182 (1984).
- Steward, R. *Science* **238**, 692-694 (1987).
- Bours, V., Villalobos, J., Burd, P. R., Kelley, K. & Siebenlist, U. *Nature* **348**, 76-80 (1990).
- Gosh, S. *et al. Cell* **62**, 1019-1029 (1990).
- Kieran, M. *et al. Cell* **62**, 1007-1018 (1990).
- Meyer, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 966-970 (1991).
- Nolan, G. P., Ghosh, S., Liou, H.-C., Tempst, P. & Baltimore, D. *Cell* **64**, 961-969 (1991).
- Ruben, S. M. *et al. Science* **251**, 1490-1493 (1991).
- Schmid, R. M., Perkins, N. D., Duckett, C. S., Andrews, P. C. & Nabel, G. J. *Nature* **352**, 733-736 (1991).
- Neri, A. *et al. Cell* **67**, 1075-1087 (1991).
- Ryseck, R.-P. *et al. Molec. cell. Biol.* **12**, 674-684 (1992).
- Lenardo, M. J. & Baltimore, D. *Cell* **58**, 227-229 (1989).
- Gilmore, T. D. *Cell* **62**, 841-843 (1990).
- Davis, N. *et al. Science* **253**, 1268-1271 (1991).
- Kerr, L. D. *et al. Genes Dev.* **5**, 1464-1476 (1991).
- Zabel, U. & Baeuerle, P. A. *Cell* **61**, 255-265 (1990).
- Ohno, H., Takimoto, G. & McKeithan, T. W. *Cell* **60**, 991-997 (1990).
- Haskill, S. *et al. Cell* **65**, 1281-1289 (1991).
- Hatada, E. N. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 2489-2493 (1992).
- Davis, L. H., Otto, E. & Bennett, V. *J. biol. Chem.* **266**, 11163-11169 (1991).
- Thompson, C. C., Brown, T. A. & McKnight, S. L. *Science* **253**, 762-768 (1991).
- Logeat, F. *et al. EMBO J.* **10**, 1827-1832 (1991).
- Inoue, J.-I., Kerr, L. D., Kakizuka, A. & Verma, I. M. *Cell* **68**, 1109-1120 (1992).
- Kawakami, K., Scheidereit, C. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4700-4704 (1988).
- Ip, Y. T., Kraut, R., Levine, M. & Rushlow, C. S. *Cell* **64**, 439-446 (1991).
- Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475-1489 (1983).
- Pognonec, P., Kato, H., Sumimoto, H., Kretschmar, M. & Roeder, R. G. *Nucleic Acids Res.* **19**, 6650 (1991).
- Smith, D. B. & Johnson, K. S. *Gene* **67**, 31-40 (1988).
- Schägger, H. & von Jagow, G. *Analyt. Biochem.* **166**, 368-379 (1987).

ACKNOWLEDGEMENTS. We are grateful for cDNA clones from C. Nüsslein-Vollhard (dorsal), I. Verma (c-rel), C. Rosen (p65), Chiron Corporation (MAD-3) and T. McKeithan (Bcl-3). We thank A. Nieters and E. Hatada for bacterial expression constructs and C. Bartsch for technical assistance. This work was supported by DFG and SFB grants to C.S.

DNA electrophoresis in microlithographic arrays

W. D. Volkmuth & R. H. Austin

Department of Physics, Princeton University, Princeton, New Jersey 08544, USA

WE have used optical microlithography to fabricate capped quasi-two-dimensional obstacle courses in SiO_2 . We report here observations using epifluorescence microscopy of the electrophoresis and length fractionation of large DNA molecules confined in arrays. Simple reptation theory, based on the work of deGennes¹, predicts that at low electric fields the electrophoretic mobility of a polymer of length L much greater than the persistence length p scales inversely with L (ref. 2). But elongation of the coil in the matrix at sufficiently strong electric fields³ results in a length-independent electrophoretic mobility^{4,5}. The application of suitably timed pulsed electric fields restores the fractionating power of gels for long molecules⁶ but the protocols of pulsed-field electrophoresis are semi-empirical because the complex and ill-understood gel matrix plays a critical role in fractionation. Microlithographically constructed obstacle arrays, with their low dimensionality, small volume and extremely reproducible topography, will make it possible to understand the motion and fractionation of large polymer molecules in complex but well characterized topologies.

An electron micrograph of a completed, microfabricated array is shown in Fig. 1a. The persistence length DNA under normal buffer conditions is $\sim 0.06 \mu\text{m}$ (ref. 7), so the DNA is confined to a vertical space slightly more than 2 persistence lengths, or one Kuhn random flight length. DNA molecules were stained by ethidium bromide and then imaged in the array by epifluorescence video microscopy^{8,9} (see legend to Fig. 1).

Figure 1b shows an image of DNA molecules moving through an array at a field strength of 1 V cm^{-1} . The DNA molecules are in focus over the entire field of view because they are confined to two dimensions. Note that the longer fragments all show substantial elongation, and that many of the longest fragments are clearly hooked by single posts. Typically, a 100 kilobase (kb) fragment hooked three times on crossing the $130 \mu\text{m}$ length field of view. The amount of elongation in this coarse array is comparable to that observed in gels, which are much finer grained¹⁰. Figure 2 displays a typical hooking sequence showing the strong lattice involvement in the dynamics. A 100 kb molecule comes in as a distorted coil, drapes around a post, hooks and then slides off. The hooked DNA molecule is actually stretched to nearly its full contour length because the observed stretched length ($30 \mu\text{m}$) is the expected contour length of a 100 kb molecule. The molecule has greater fluorescence intensity at its ends furthest from the post (Fig. 3). This is presumably due to the randomization of the ends of the molecule where the internal tension is very small¹¹.

Our arrays correspond to 'unphysical' agarose gels. The effective pore size of $1.0 \mu\text{m}$ corresponds roughly to a physically unstable 0.05% agarose gel¹². Such a large pore spacing is comparable to the $1.5 \mu\text{m}$ radius of gyration of the largest DNA molecules. The amount of elongation and hooking that occurs in the arrays is not explained by simple models in such an open lattice. Of course, the applied electric field of 1 V cm^{-1} is not responsible for the alignment of the polymer in the absence of the lattice, as is quantitatively seen by evaluation of the dimensionless number $\kappa_a = -QE p / kT$ (where k is Boltzmann's constant, T is the absolute temperature and Q is the net charge of a persistence length)¹³. If $\kappa_a < 1$ then the entropy of the polymeric chain dominates and the coil is random, whereas if $\kappa_a > 1$, the potential energy gained in translation of the persistence length along the field direction dominates and the chain is elongated. Because the applied electric field E is 1 V cm^{-1}

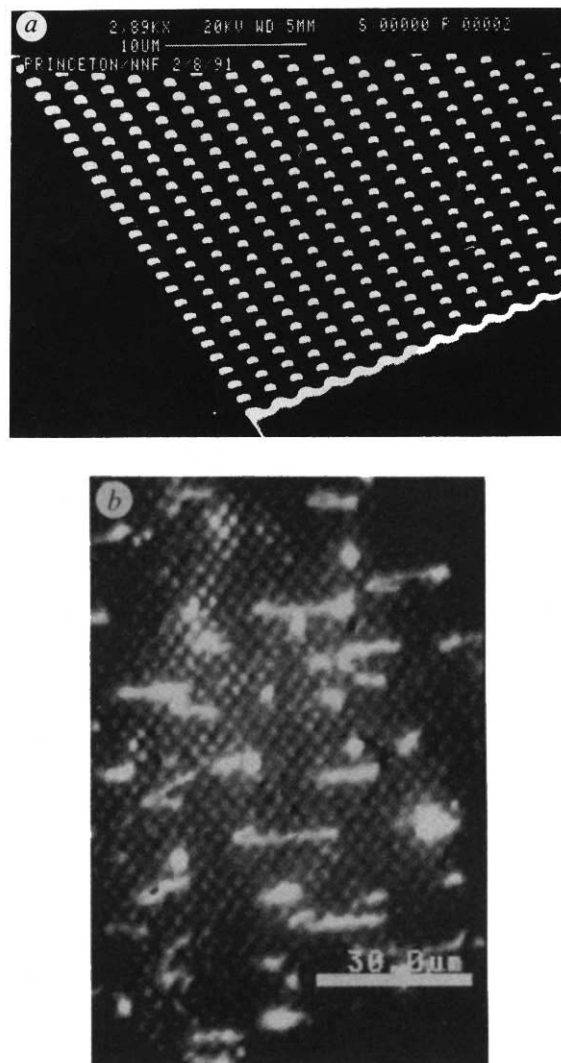


FIG. 1 a, An electron micrograph of a corner of an array. The micrograph shows the $0.15 \mu\text{m}$ -high posts, diameter $1.0 \mu\text{m}$, and centre-to-centre spacing of $2.0 \mu\text{m}$. The posts in this array form linear rows in the direction of the electric field, whereas in the epifluorescence pictures shown we used a staggered array of posts. b, An epifluorescence view of DNA molecules moving through the array. The DNA molecules are moving to the right. We fabricated arrays on a standard 3 inch diameter silicon wafer. A $0.5 \mu\text{m}$ SiO_2 layer was grown into the surface of the array for optical emission measurements. The arrays consisted of a $2.7 \text{ mm} \times 2.7 \text{ mm}$ rectangle containing 2 million posts. The etch depth of the pattern was variable. In the experiments we describe here the posts are $0.15 \mu\text{m}$ high, $1.0 \mu\text{m}$ in diameter, and $2.0 \mu\text{m}$ centre-to-centre. We made one set of arrays with the posts on a square lattice, and another set also on a square lattice but with the lattice rotated 45° . A flat 'loading' area etched to the same depth as the posts from the surface was placed at the two ends of the array to facilitate loading the DNA-containing solution into the array and can be seen on the left portion of the micrograph. A Pyrex coverslip (Gould Precision Optics, Port Crane, New York) was fused to the tops of the posts using field-assisted silicon-glass fusion⁹. Tests with fluorescent dye revealed that all the post tops were sealed by the fused pyrex coverslip. The samples of DNA are highly purified *Micrococcus luteus* (Sigma Chemical Ultra-pure) originally of uniform length 100 kb ($\sim 30 \mu\text{m}$ contour length). The DNA fragments were deliberately sheared in a pipette to provide a distribution of lengths. Ethidium bromide at a concentration of $50 \mu\text{g ml}^{-1}$ was used for these high-contrast figures. Some of the longer molecules can be seen hooking around posts. A Nikon Optiphot-2 microscope was used with fluorescence excitation at the mercury line of 546 nm. A Nikon $\times 60$, NA 1.4 oil immersion objective was used. A microchannel plate image intensifier ($\times 10,000$ optical gain) was used followed by a Burle CCD camera. Video signals were stored on a NEC PC-VCR and images were 'grabbed' by a Cortex Framegrabber for processing on a Silicon Graphic Iris computer using code developed by the Graphics Laboratory at Princeton University.

FIG. 2 A sequence of false-colour images, 'grabbed' at ~ 5 -s time intervals, showing (from top to bottom): *a*, 100 kb fragment in a distorted coil configuration approaching from the right; *b*, hooking on a post; *c*, extended on a post with the characteristic enhanced brightness on the ends of the molecule; and *d*, sliding from the post.

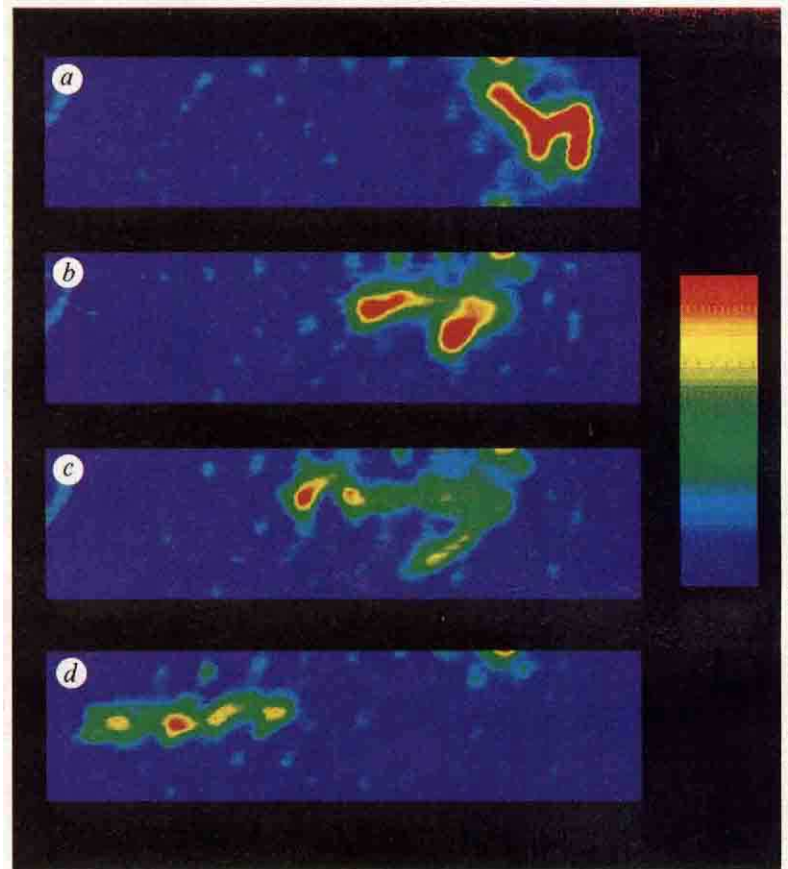
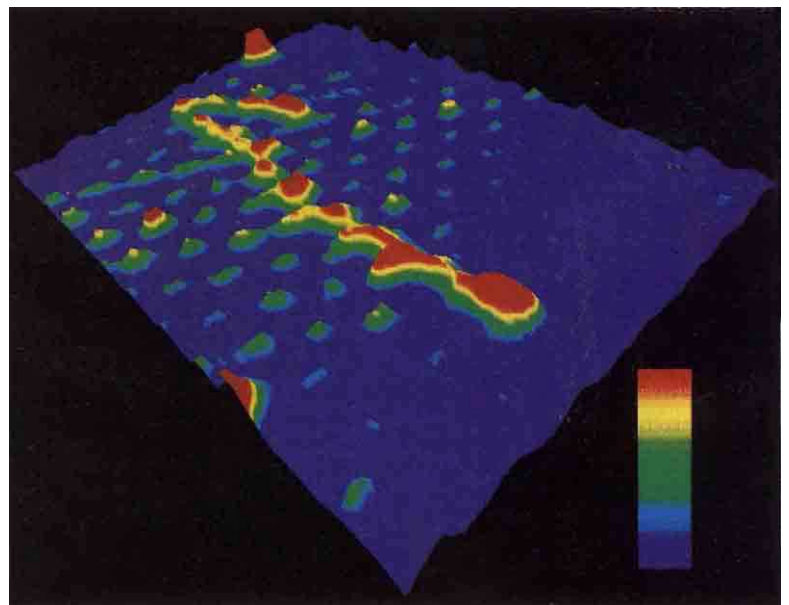


FIG. 3 A three-dimensional false-colour image of the hooked DNA molecule, digitally zoomed to $\times 4$ the magnification in Fig. 2. The intensity of fluorescence light is represented both by z-axis height and by colour. The posts are visible as small hills, and a $30\text{ }\mu\text{m}$ long DNA molecule is visible as a long U-shaped ridge. As discussed in the text, the unstrained ends of the molecule are entropically disordered and show greater intensity, whereas the inner parts of the molecule stretch to the single helix thickness and have a uniform intensity.



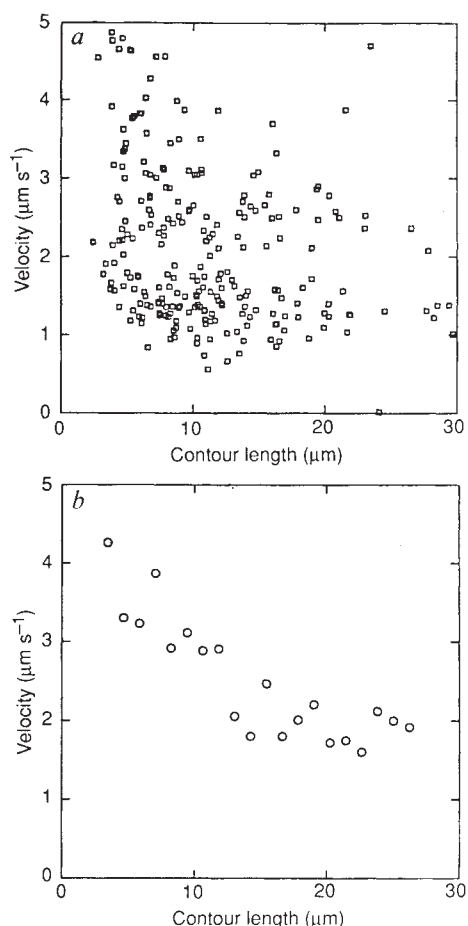


FIG. 4 *a*, A scatter-plot of the observed velocities of different sizes DNA molecules as a function of length. *b*, A filtered plot of *a* done by binning the data and determining the average velocity within the bin. Because the optical response of the camera we have is roughly linear in photon flux, we assumed that the light at any spot in the image was proportional to the mass of the stained DNA. Therefore we were able to calculate the motion of the centre-of-mass of the molecules as a function of both time and length. Our computer scanned successive frames of a background-subtracted sequence of DNA molecules moving through the array. The code determined the length of each molecule by integrating the total light emission from a given connected molecule. Each molecule had to be traced for at least six successive frames, spaced 0.5 s apart, to be counted as a legitimate case. The large amount of scatter in the unsmoothed plot is predominantly due to the episodic hooking of the DNA molecules in our small $130 \times 100 \mu\text{m}$ field of view. The binning length for this plot was set at $5 \mu\text{m}$.

and the net charge per unit length should be $0.1 e^-$ per $3.4 \text{ \AA}$ (ref. 14) we have $\kappa_a = 0.005$.

This small value for κ_a is not consistent with the observations of the elongation of the hooked molecules. The characteristic distance L' over which this thinning from random coil at the ends to fully stretched polymer has been calculated by Schurr and Smith to be $L' \sim 2p/\kappa_a$ (ref. 11). In our case, a $\kappa_a = 0.005$ gives a value for L' of $\sim 30 \mu\text{m}$ or 100 kb, whereas the measurements reveal a much smaller L' of $\sim 5 \text{ kb}$ (about $1.5 \mu\text{m}$). This may be due to electroosmosis¹⁵ in our relatively open array. The hydrodynamic drag from electroosmosis causes a larger force per unit length on the molecules, effectively increasing E . The influence of two-dimensional statistics in the array may also play a significant role in the dynamics of the polymers.

Finally, we consider the fractionating ability of these arrays. Figure 4 shows the mobility of DNA observed in our array as a function of length. The binned data in Fig. 4 indicate that up to a length of $\sim 100 \text{ kb}$, the array is capable of length fractionation in a DC field. This is at the limits of fractionation for

conventional DC electrophoresis in agarose gels¹⁶, giving hope that our arrays can extend the present limits of gel electrophoresis. Noolandi first suggested the importance of hooking in gels¹⁷ and Song and Maestre¹⁸ have done preliminary analysis of unhooking times. We believe that fractionation in our arrays occurs due to dispersion in the time it takes a hooked polymer to thermally free itself from a post. We hope that by judicious application of careful array design and attention to theoretical guidance, optimization of length fractionation will be possible. $\square$

Received 19 March; accepted 10 June 1992.

- DeGennes, P. *Scaling Concepts in Polymer Physics* 4th edn, 223-233 (Cornell Univ. Press, Ithaca, New York, 1991).
- Lerman, L. S. & Frisch, H. L. *Biopolymers* **21**, 995-997 (1982).
- Smith, S. B., Aldridge, P. K. & Callis, J. B. *Science* **243**, 203-206 (1989).
- Hervet, H. & Bean, C. P. *Biopolymers* **26**, 727-742 (1987).
- Lumkin, O. J., Dejaridin, P. & Zimm, B. H. *Biopolymers* **24**, 1573-1593 (1985).
- Schwartz, D. C. & Cantor, C. R. *Cell* **37**, 67-75 (1984).
- Hagerman, P. J. *A. Rev. Biophys. biophys. Chem.* **17**, 265-286 (1988).
- Morikawa, K. & Yanagida, M. *J. Biochem.* **89**, 693-696 (1981).
- Bustamante, C. A. *Rev. Biophys. biophys. Chem.* **20**, 415-446 (1991).
- Schwartz, D. C. & Koval, M. *Nature* **338**, 520-522 (1989).
- Schurr, J. M. & Smith, S. B. *Biopolymers* **29**, 1161-1165 (1990).
- Holmes, D. & Stellwagen, N. *Electrophoresis* **11**, 5-15 (1990).
- Deutsch, J. M. & Madden, T. L. *J. chem. Phys.* **90**, 2476-2485 (1989).
- Manning, G. S. *J. chem. Physics* **51**, 924-933 (1969).
- Kozak, M. W. & Davis, E. J. *Langmuir* **6**, 1585-1590 (1990).
- Serwer, P. *Electrophoresis* **10**, 327-331 (1989).
- Lalande, M. *et al. Nucleic Acids Res.* **16**, 5427-5437 (1988).
- Song, L. & Maestre, M. F. *J. biomolec. Struct. Stereodynamics* **9**, 87-99 (1991).
- Wallis, G. & Pomerantz, D. I. *J. appl. Phys.* **40**, 3946-3949 (1969).

ACKNOWLEDGEMENTS. We thank the National Nanofabrication Facility for assisting in the construction of the arrays, in particular H. Craighead, L. Rathbun and G. Pugh, M. C. Wu and M. Ivker for some initial array fabrications, W. W. Webb, D. Piston, K. McDonald, J. Goodhouse, M. Rose and E. Fairfield for encouragement and advice on optical microscopy, P. Burboeck of Micron Optics for loans of optical equipment, K. Alexander for computer graphics, E. C. Cox for advice and encouragement and S. Leibler for theoretical insights. This work was supported by the National Science Foundation and the Office of Naval Research.

nature is available in microform

University Microfilms

International reproduces this publication
in microform; microfiche and 16mm or 35mm film.

For information about this publication or any of the
more than 13,000 titles we offer, complete and mail
the coupon to: University Microfilms International,
300 North Zeeb Road, Ann Arbor, MI 48106.

Call us toll-free for an immediate response;

800-521-3044.

Or call in Michigan, Alaska and Hawaii:

313-716-4700.

☐ Please send information about these titles:

.....

Name

Company / Institution

.....

Address

City

State Zip

Phone

Automated real-time immunoassay of biomolecules

N. B. Afeyan, N. F. Gordon and F. E. Regnier

ImmunoDetection is a novel technique combining perfusion chromatography technology with antibodies to perform the steps of an immunoassay in a flow-through column format. Sensitive and precise measurements are performed in seconds to minutes using automated liquid chromatography instrumentation.

ANALYTICAL techniques used to measure the concentration of a target substance (analyte) in a mixture often involve a separation step followed by detection (for example, electrophoresis, chromatography). Liquid chromatography (LC) is one such technique in which the separation is accomplished by differential migration of species through a packed column of particles. Detection of separated components is usually accomplished by measuring UV absorbance. When used to measure the amount of one substance in a mixture, LC is at best an indirect measurement, which relies on the reproducibility of the elution time and the degree of separation of the target from other solutes.

A more direct measurement of the concentration of one substance can be achieved by the technique of immunoassay. In this case, the analyte is separated from other components of a mixture by binding specifically, and with high affinity, to an antibody or other binding molecule. Immunoassays are typically performed in a microtitre-well format and are usually slow (several hours) and labour intensive. ImmunoDetection (ID)¹ is a novel approach that combines the advantages of both LC and immunoassay to perform fully automated, real-time measurements of analyte concentration. Previous approaches to the use of affinity chromatography provide the basis of this technique²⁻⁴. The article describes the principles of ID, its performance characteristics, aspects of automation and novel applications.

ImmunoDetection

The ImmunoDetection technique is based on the use of recently developed, very high speed LC materials⁵. These perfusion chromatography media, when used in combination with antibodies or other ligands that can bind to a complementary biomolecule, allow complete analyte capture in seconds. The ID sensing device consists of three elements: the immobilized ligand, a flow-through particle and a flow-through cartridge. A schematic representation showing the flow-through cartridge at increasing magnification is shown in Fig. 1.

The flow-through cartridge (column) format allows ID assays to be performed using conventional HPLC or FPLC instruments. Similar to LC columns, the cartridge is intended for reuse hundreds of times. Additionally, the automation and data handling components of modern LC instrumentation

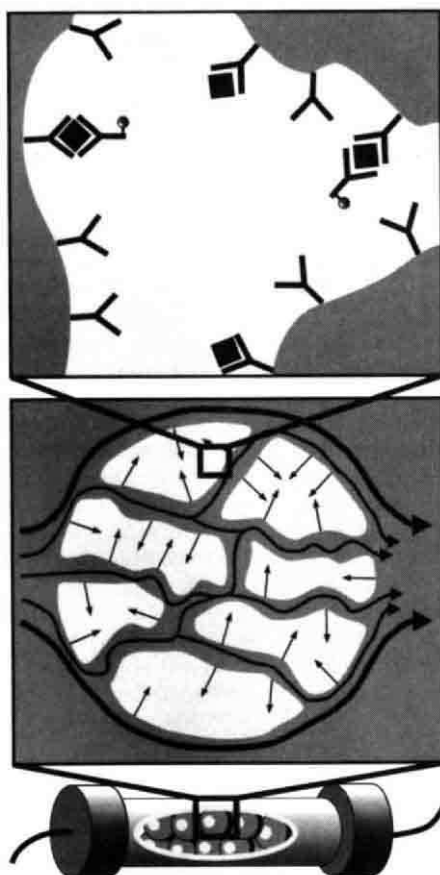


FIG. 1 Schematic of ID sensor elements: flow-through cartridge (column), flow-through bead (short arrows depicting diffusion and long arrows depicting connective flow), pore space showing immobilized antibodies, captured antigens and secondary, labelled antibodies.

ments are directly applicable to ID assays.

With the ID assay format, the antibodies are immobilized at the pore surface contained within the pores of the particle. Essentially, the pores function much like single wells contained within a microtitre plate. The target biomolecules bind to the antibodies with high affinity and specificity. By capturing the target biomolecules on the pore surface, they can easily be separated away from all other sample components.

The flow-through beads, the basis of perfusion chromatography, provide a large amount of readily accessible surface area ($>20 \text{ m}^2 \text{ g}^{-1}$), analogous to a device containing thousands of interconnected microtitre wells. Solutes are rapidly convected into the

interior of the particle, creating a sub-micrometre diffusion path length to the smaller pores where the bulk of the antibody is immobilized. Consequently, target biomolecules in the flowing stream are rapidly brought into intimate contact with an excess of immobilized antibody, an arrangement that promotes rapid target capture. Long incubation steps are eliminated as cartridge residence times of less than one second are sufficient for complete capture⁶.

The ID assay is performed in a flow-through cartridge, housing millions of beads. This device does not restrict the amount of sample used to a fixed volume. If the target biomolecule concentration is low, a large volume of sample is utilized, the target is concentrated within the cartridge and then easily detected. Conversely, if the target biomolecule concentration in the sample is high, a small volume of sample is utilized, obviating the need for serial dilutions of the sample. The cartridge provides a dynamic assay range spanning four orders of magnitude, easily accommodating wide ranges of target biomolecule concentrations.

The flow-through cartridge format also provides for highly effective washing and removal of all nontarget components of the sample, minimizing background interference and enhancing analytical precision. Once the nontarget solutes are washed off the cartridge, the captured analyte is eluted and typically detected directly by UV absorbance. The immobilized antibody is regenerated in loading buffer and can be reused for hundreds of cycles.

Antibody, or in general any binding molecule, can be immobilized to the pore surface inside the particles using one of three approaches. In the first approach, antibody is captured on an immobilized Protein-A or Protein-G surface, enabling the use of crude antibody preparations without compromising antibody coupling efficiency. Subsequently, antibody is covalently anchored with a bifunctional crosslinking agent (dimethyl pimelimidate). This chemistry orients the antibody with the Fab regions exposed, providing for a high antibody specific activity. In addition, the immobilization does not require any excess antibody to drive the reaction. The second approach is to attach the antibody directly and covalently to the particle surface. A variety of activation chemistries can be used including aldehyde, epoxide and hydrazide

derivatizations. Finally, an avidin-activated surface can be used to immobilize selectively biotinylated antibodies.

Assay performance

An ID assay was developed for the direct measurement of human serum albumin (HSA). A cartridge (2.1-mm diameter $\times$ 30-mm length) with covalently immobilized anti-HSA was used. A typical assay run took one minute to complete and was sensitive down to 100–500 ng of analyte. Figure 2 shows the calibration curve generated using HSA as standard. The broad dynamic range of this assay, from 100 ng to 100 μ g, is due to the cartridge acting as an analyte concentrator. Detection of HSA was performed after elution with acid (5 mM HCl) by measuring UV absorbance (220 nm). This on/off assay was run 200 consecutive times using the same ID cartridge with human serum as the sample.

Although UV absorbance is the simplest detection mode, more sensitive ID assays can be performed by using the common signal amplification techniques that use enzymes, fluorescence or chemiluminescence. For these cases, competitive or sandwich immunoassay formats are adapted to the flow-through cartridge format. For example, an enzyme-linked sandwich assay is performed using the following method. Primary antibody is first immobilized on the cartridge. Sample is then directed through the cartridge and the target biomolecule is captured. Subsequently, an enzyme-tagged secondary antibody is injected and forms a sandwich with the primary antibody and captured target. Substrate solution is then pumped over the cartridge and the formation of a product from the enzymatic turnover of the substrate is monitored using a visible light detector. The minimum detection limit of this approach is on the order of 10 pg as shown in the calibration curve in Fig. 2. This detection amplification technique is complementary to the on/off technique, which is useful at higher analyte concentrations.

ImmunoDetection assays offer several potential advantages relative to conventional, microtitre plate assays. (1) The ID assays operate with much higher analytical precision, with coefficients of variation of ap-

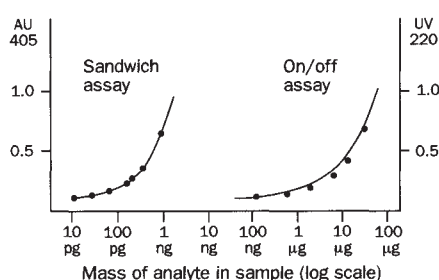


FIG. 2 Dynamic range for ID assay formats. A calibration plot for human serum albumin (HSA) detection using an anti-HSA cartridge with an on/off assay format and UV detection at 220 nm (right). A similar plot using an enzyme-amplified sandwich format with alkaline phosphatase, p-nitrophenyl phosphate (pNPP) as a substrate and detection at 405 nm (left). Analyte mass plotted on logarithmic scale; same data plotted on a linear scale yields straight lines with regression coefficients >0.99 .

proximately 1%. (2) The large dynamic range of the ID cartridge as well as the ability to utilize small-volume samples (for example, 1 μ l) minimizes and often eliminates the need for sample dilution. Manual pipetting of samples, to dilute them to the low concentrations required for conventional plate immunoassays, is a well recognized source of experimental error. (3) The flow-through assay format provides an efficient means of extensively washing away nontarget solutes and excess labelled reagents that otherwise produce large and variable background signals. (4) The ID cartridge device is also directly calibrated such that standards and unknowns are assayed under identical conditions. This direct calibration contrasts with conventional plate assays where standards and samples are run in different wells. Small well-to-well differences introduce assay variability and contribute to the large coefficients of variation typically encountered with plate assays (10–20%).

ImmunoDetection assays are fast, with typical run times on the order of 1–3 min. These assay times make it feasible to perform single assays in a sequential mode, versus the parallel processing of multiple samples associated with conventional plate immunoassays. Moreover, the ability to perform rapid assays shortens the time

frame required to develop and validate the assay. Typically, a new UV-based assay can be developed, validated and operational in a few hours.

Assay automation

The ID assays are designed to be used on HPLC instruments, which are in turn designed for automation. As is typical with HPLC, large numbers of samples can be loaded in autosampler trays and processed in an unattended mode. The precise sample and reagent delivery characteristics of HPLC instrumentation also add to the analytical precision of an ID assay. Additionally, when performing conventional immunoassays, one spends considerable effort running many different sample dilutions searching for the dilution that yields a result within the narrow dynamic range of the assay. Conventional immunoassays typically have a narrow dynamic range due to the small quantity of antibody that is immobilized on the surface of the microtitre well. In contrast, the ID assay can accommodate a larger range of analyte concentrations, greatly reducing the number of sample dilutions that need to be run per sample. Finally, another important aspect of automation is the reuse of the ID cartridges. Unlike conventional immunoassays, where reagents are used once, an ID cartridge can be reused hundreds of times. □

N. B. Afeyan and N. F. Gordon are at PerSeptive Biosystems, Inc., University Park at MIT, 38 Sidney Street, Suite 100, Cambridge, Massachusetts 02139, USA and F. E. Regnier is at Purdue University, West Lafayette, Indiana 47907, USA. For more information, fill in reader service number 100.

1. Regnier, F. E., Xu, C., deFrutos, M., Birck-Wilson, E., Dorvel, B. & Afeyan, N. B. 11th Int. Symp. HPLC Proteins, Peptides and Polynucleotides (Washington, DC, USA, 1991).
2. Ohlson, S., Hansson, L., Larsson, P. O. & Mosbach, K. *FEBS Lett.* **93**, 5–8 (1978).
3. De Alwis, W. U. & Wilson, G. S. *Analyt. Chem.* **57**, 2754–2756 (1985).
4. Hage, D. S. & Walters, R. R. *J. Chromatogr.* **386**, 37–49 (1987).
5. Afeyan, N. B. et al. *J. Chromatogr.* **519**, 1–29 (1990).
6. Regnier, F. E. *Nature* **350**, 634–635 (1991).
7. Afeyan, N. B., Gordon, N. F., Khachatrian, L. & Regnier, F. E. *HPLC 92* (Baltimore, Maryland, USA, 1992).

ADVERTISEMENTS

Synthetic GENES
GGG GAA AAC GAA TCC
GGT GAA AAT GAA TCC
GGT GAA AAT GAA TCC

TIB
MOLBIOL

for further information
FAX +49 30 7127734

Reader Service No.18

Professional Custom Services

DNA SYNTHESIS
PROTEIN SEQUENCING
PEPTIDE SYNTHESIS

Rapid Service • Guaranteed Quality • Ample Quantity
Competitive Prices • Free Quotations
World-wide Delivery

V.G. Services 1200 Wonderland Rd S. Bldg #7 Unit #2
London, Ontario Canada N6L 1A8
CALL/FAX: (519) 652-3758 ALT. FAX: (519) 652 0053

Reader Service No. 21

\$3.50/Base
Custom DNA
Synthesis

No Set-up Fee,
Shipped in 24 hours.

1-800-788-4362
1-800-786-4362 Fax

KEYSTONE
Laboratories, Inc.

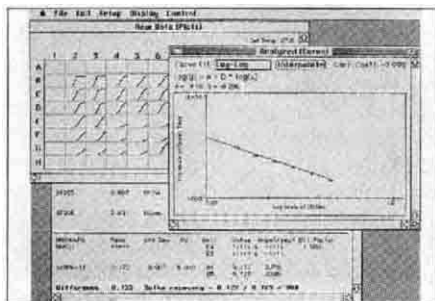
Reader Service No.33

Laboratory aids for the immunologist

Featured this week — a compendium of immunological reagents, microplate analysis software and an *in vitro* cell culture system for the production of monoclonal antibodies, which provides an alternative to ascites production.

If wading through catalogues to find out which companies sell which reagents is not your cup of tea, then the Linscott's **directory of immunological and biological reagents** may be the answer (*Reader Service No. 101*). The new seventh edition contains information on where to buy 8,500 monoclonal antibodies, 12,500 conventional antisera and conjugates and 18,000 other reagents including immunoassay kits, enzymes, purified serum fractions, blood products, recombinant DNA reagents, hormones and peptides, separation media, tissues, organs, microbial antigens and cultures, cell lines (including many hybridomas) and unusual laboratory animals. International reference standards from the World Health Organization and many different reagents available through the National Institutes of Health are also included. To help combat obsolescence, supplementary updates to the directory appear every six months for a year and a half (when the next edition will be published). These supplements are included in the \$70 (US) purchase price (US, Canada and Mexico), \$80 (US) (Central and South America), other countries (price on request).

Molecular Devices introduces a new generation of **SOFTmax microplate analysis software** for both Windows and Macintosh platforms (*Reader Service No. 102*). The software, which is designed to complement THERMOMax, UVmax, Vmax and Emax microplate readers, incorporates the full range of analytical and kinetic capabilities of the current version but with many new analysis and reporting features. In addition to performing ELISAs, microplate assay capabilities such as cytoproliferation, cytotoxicity, neutrophil activation, enzyme kinetic studies and endotoxin detection can now be performed using a mouse-driven interface with full 'windowing' capabilities. A new pull-down assay menu allows users to recall frequently used assay protocols with one operation. Spike recovery is now automatically calculated and reported. The

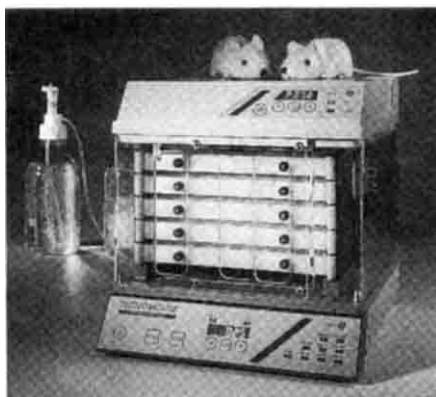


Dual platform microplate software.

company says that an improved four parameter curve-fitting capability is provided. Onset time calculation allows clot formation and dissolution assays to be quantified. SOFTmax allows for onscreen display of concentration, which is intended to make it easy for users to verify the accurate entry of dilution values.

■ *In vitro* cell culture systems

Tecnomouse is the new ***in vitro* cell culture system** from Northumbria Biologicals that is designed to provide an economical and automated alternative to ascites production of monoclonal antibodies (*Reader Service No. 103*). The system consists of five



Tecnomouse — the ascites alternative.

bioreactor cassettes, each equivalent in production terms to 20 laboratory mice, the company claims. By automating the procedure, skilled personnel can be released to more productive work.

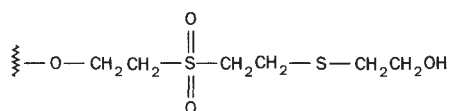
Micro Mouse is a hollow-fibre cell culture system from UniSyn Technologies designed especially for the **laboratory-scale production of human and murine monoclonal antibodies** and recombinant proteins (*Reader Service No. 104*). According to UniSyn, it can replace up to 15 mice in monoclonal antibody production. The presterilized flowpath contains a two-litre media reservoir and septa sample ports for easy maintenance and recovery of the product. The mini bioreactor utilized in the Micro Mouse is available in three different membrane pore sizes.

■ What a bind!

Protein G-prime is a new **antibody binding protein** from Aaston that binds IgG-type antibodies specifically and exclusively (*Reader Service No. 105*). Binding takes place at the Fc region of the antibody, leaving the antibody activity unaffected. This new binding protein is based on native

Protein G but has, Aaston says, undergone molecular engineering modifications to permit efficient and simple immobilization and to prepare the protein for easy conjugation to marker molecules. Molecular engineering techniques were used to delete regions of the protein that normally bind to the Fab region and albumin. Aaston says that Protein G-prime has the same high affinity for IgG antibodies as its native precursor. It shows none of the binding to albumin that limits the specificity of native Protein G and does not display the low affinity binding to the Fab region of antibodies that the native protein exhibits. In addition, the company says that Protein G-prime shows no loss of activity even after heat treatment for 20 min at 100 °C. It also has unaltered activity following exposure to pH 1.5–11.

T-Gel thiophilic matrix is a new **Protein-A alternative** from BioAffinity Systems (*Reader Service No. 106*). The patented invention of Jerker Porath, T-Gel has a capacity of 10–20 mg ml⁻¹ for IgG from a wide variety of sources (goat, rabbit, rat, bovine, murine and human), allowing for rapid and flexible purification of antibodies. Typically, yields and purity are >90 per cent, BioAffinity says. Binding and washing is accomplished with 0.5 M K₂SO₄ containing 50 mM phosphate, pH 8 and



Thiophilic ligand on T-Gel.

elution with 50 mM phosphate, pH 8. T-Gel can be sterilized and reused many times. In addition, the crosslinked agarose support bearing the thiophilic ligand has good flow-rate characteristics and can withstand moderate pressure (up to 25 p.s.i.) without collapse, the company claims. BioAffinity is offering a T-Gel trial kit, which includes a prepacked plastic mini-column (with frits, caps and bottom plugs) containing 5-ml bed volume of T-Gel, a starting supply of binding buffer and ACS-grade K₂SO₄, together with instructions. Other pack sizes of T-Gel are available.

IBI has extended its FLAG product line to include the anti-FLAG M2 affinity gel for the **single-step purification of recombinant FLAG fusion proteins** (*Reader Service No. 107*). Using the anti-FLAG M2 affinity gel, FLAG fusion proteins can be purified by elution with glycine or by competition



One-step purification of FLAG fusion proteins.

with free FLAG peptides. IBI says that the affinity column can be used for the purification of fusion proteins that contain an amino terminal, methionine amino terminal, internal or carboxy terminal FLAG marker octapeptide. It has been used successfully for the purification of FLAG fusion proteins in animal, insect, yeast and bacterial cells. The DNA coding sequence for the low-molecular-weight FLAG marker is only 24 bases long, which can be easily synthesized and inserted into the expression system of choice or, alternatively, used directly in IBI's bacterial expression system. The column consists of the anti-FLAG M2 IgG₁ monoclonal antibody coupled to an agarose matrix via a hydrazide linkage. It is available in 1-, 5-, 10- and 25-ml quantities with or without FLAG peptides for competitive elution applications.

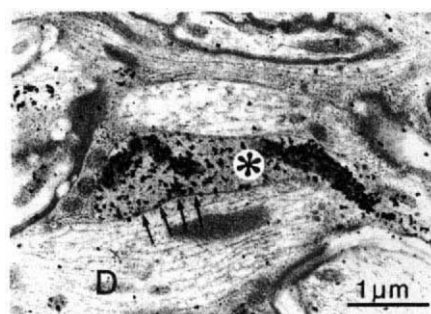
■ Assorted immunoreagents

Strattech Scientific now has available two new **immunodiagnostic coating solutions and a blocking agent** (*Reader Service No. 108*). Panacoat will immobilize antibodies, antigens, enzymes and other proteins on to plastic surfaces such as tubes, beads, microtitre plates and particles. Strattech says that Panacoat provides thermal stability to the coated surface and the coating achieved is uniform with minimal batch-to-batch variations. Most types of plastic can be coated—even Nylon. The second immunodiagnostic coating solution, Chemicoat, provides covalent linkage as well as hydrophobic bonding of antibodies, antigens and other proteins to polystyrenes and polypropylenes and some other types of plastic. Completing the line-up is Panablock, which can be used for blocking unbound sites when preparing solid-phase systems. When used in very low ionic concentrations, Strattech says that Panablock will block with minimal background and high signal in immunological reactions.

Gibco-BRL is distributing a nonionic medium, Nycodenz, which is designed to make lightwork of the **isolation of human mononuclear cells** (*Reader Service No. 109*). Nycodenz is a solution of Nycodenz and sodium chloride that can be mixed directly with whole blood. No layering of the sample is necessary and

the recovery and viability of the harvested mononuclear cells is equal to or better than conventional density media (and the erythrocytes are not aggregated), says Gibco-BRL. This one-step method for isolation is designed to avoid the artifactual banding of erythrocytes at the sample/medium interface. In addition, the absence of carbohydrates, such as Ficoll or dextrans, will leave the cell receptors unaffected.

Nanoprobes has developed a new **gold immunoprobe** that is designed to improve significantly the results achieved with electron microscopy (*Reader Service No. 110*). Nanogold is a 1.4-nm gold compound that is covalently attached to the Fab' fragment (or IgG). A brief exposure to Nanoprobes HQ Silver enlarges the small gold particles to ~10 nm so that they may be easily identified, Nanoprobes says. The accompanying electron micrograph shows GABA-containing terminals (asterisks) forming symmetrical synaptic specializations (arrows) with dendrites (D) in the thoracic spinal cord of the rat. The presence of



Immunostaining for electron microscopy (courtesy of S. Bacon, Oxford University, UK).

GABA within the terminals was revealed by post-embedding staining using GABA antiserum (4 °C, 18 h), Nanogold (goat anti-rabbit IgG, room temperature, 90 min) and intensified for 6 min using HQ Silver (Nanoprobes). Tissue was counterstained with lead citrate and the embedding resin was Durcupan.

■ ELISAs

British Bio-technology has introduced a new **ELISA kit to measure soluble ICAM-1** (intercellular adhesion molecule) in sera, plasma and tissue culture supernatants (*Reader Service No. 111*). ICAM-1, a member of the immunoglobulin superfamily and a cell-surface glycoprotein, is involved in adhesion of leucocytes to endothelium through LFA-1 and Mac-1 ligands (which are members of the β_2 (CD18) family of integrins). British Bio-technology says that initial evidence that ICAM-1 may exist in a soluble form comes from the observation that lymphocytes and Jy cells release a soluble form of ICAM-1 when cultured *in vitro*. Subsequently, blotting techniques and ELISA methods have shown



Soluble ICAM-1 ELISA kit.

that the soluble version of ICAM-1 appears to be present in serum of normal individuals and it is thought that soluble ICAM-1 may have a role in the immune disease process — diseases such as rheumatoid arthritis, diabetes, systemic lupus erythematosus, metastatic cancer, acute urolithiasis and cardiac, liver and renal transplant rejection. The biological significance of circulating soluble ICAM is not clear. Serum levels may be explained by the release of soluble ICAM-1 from cells after minor inflammatory episodes, which the body suffers on a daily basis. British Bio-technology says that raised ICAM-1 levels may, therefore, provide an indication of increased inflammation or immune activity.

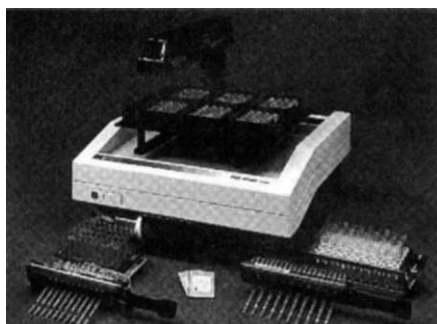
Innotest hGM-CSF is a new microplate-based ELISA kit from Innogenetics for the **quantitative determination of human granulocyte macrophage-colony stimulating factor (hGM-CSF)** in serum, plasma and cell culture supernatants (*Reader Service No. 112*). The assay can be used to quantify biologically active, native and recombinant hGM-CSF in samples during a single test run. It has a detection limit of 8 pg ml⁻¹ a measuring range of 0–320 pg ml⁻¹ and a recovery of 92 per cent. Cross reactivity with human IL-4, human IL-5 and human macrophage-colony stimulating factor has been tested and found to be lower than 0.002 per cent, the company says. Ready-to-use standards are provided.

For researchers involved in arteriosclerosis research, Boehringer Mannheim offers the **lipoprotein (a) ELISA** for the quantitative determination of lipoprotein (a) (*Reader Service No. 113*). This sandwich-type enzyme immunoassay works as follows: an anti-lipoprotein (a) polyclonal capture antibody is absorptively bound to the wells of a microtitre plate and coated with a blocking reagent; the wells are then incubated with the standard or sample; anti-lipoprotein (a) monoclonal antibody-peroxidase conjugate is added and binds to the captured lipoprotein (a); the addition of the peroxidase substrate, ABTS, turns the

wells to green in proportion to their lipoprotein (a) content. Absorbance values (at 405 nm) obtained from the standards are used to plot a linear curve against which the samples are read.

■ Handy hardware

Eliminate the tedious manual washing step from your radioimmunoassay (RIA) workload with the new line of EXEC-WASH automated washing systems for every RIA application from Source Scientific Systems (Reader Service No. 114). Researchers can choose from nine different EXEC-WASH RIA configurations with applications specific for bead, coated tube, star tube, bead tray and microwell cup formats. This allows researchers to match a gamma-counting or reagent kit racking system to an EXEC-WASH RIA configuration



Automated RIA washing alternatives.

or select a generic washing system for all-purpose use. By automating the washing step in RIA, the EXEC-WASH solid-phase washing system ensures consistent performance, minimizes labour and reduces exposure to biohazards, Source Scientific says. In the fully self-contained system, sample tubes are washed, transported and analysed in the instrument's gamma-counting racks. The EXEC-WASH's system design minimizes sample-tube handling and potential mix-ups of tubes. Users can create custom RIA wash procedures using the company's Protocol Designer software.

Immunelectrics introduces new Miniblotters for western blot screening of individual miniblots (Reader Service No. 115). Two models are available: the MN28SL, which has 28 channels with each channel holding as little as 50 µl and the MN20SL, which has 20 channels and is designed to optimize high-sensitivity detection of low-titre or low-affinity antibodies. The channel capacity of this model is 350 µl. Intact preblotted membranes are mounted directly in the Miniblitter, which eliminates the need to cut the membrane into strips as with conventional western blotting methods. The multiple channels function as separate miniature incubation trays. Immunelectrics says that cross-channel contamination is eliminated because of the patented sealing system. A washing manifold plugs into the Miniblitter, allowing all the channels to be rinsed simultaneously in seconds. The Miniblitter units are modular and compatible with other instrumentation from Immunelectrics. Custom gel combs are available, enabling users to match gel lanes with Miniblitter channels. In this way, multiple samples on a minigel can be blotted against different antibodies. The units are designed for serum and monoclonal antibody screening applications. Both the MN28SL and MN20SL cost \$395 (US) each; \$535 (US) each if a manifold is included.



Miniblotters for western blot screening.

Antibodies Cambridge Research Biochemicals has added four new products to its range of oncoprotein antibodies (Reader Service No. 116). The line-up includes two rabbit polyclonal antibodies to p56lck recognising different epitopes (p56lck is a tyrosine protein kinase that forms complexes with CD4 and CD8 in T-lymphocytes and NK cells). CRB also offers a rabbit polyclonal to *trk* (*trk* belongs to the family of tyrosine protein kinase receptor molecules). The *trk* protooncogene product has been shown to be the high-affinity nerve growth factor receptor. Completing the line-up is a rabbit polyclonal antibody to the retinoblastoma gene product (the retinoblastoma gene product is a nuclear phosphoprotein that is thought to regulate cell proliferation). Mutations in the retinoblastoma gene are found frequently in human sarcomas, lung, bladder and breast cancers and are the molecular basis for hereditary predisposition to retinoblastoma.

■ Antibodies

A determination of serum calcitonin is useful for the clinical diagnosis of early-stage medullary thyroid carcinoma and for monitoring calcitonin therapy. Mouse monoclonal antibodies to human calcitonin are now available from Crystal Chem as a purified IgG preparation (Reader Service No. 117). A synthetic peptide of calcitonin 1-32 was used as an immunogen for the preparation of the monoclonal antibodies. The company is also offering recombinant alpha-fetoprotein domain 1 (r-AFP domain 1) prepared from genetically engineered *Escherichia coli*. Immunologically, r-AFP domain 1 reacts with the antibody against human AFP but it does not react with the antibody against human albumin.

These notes are compiled by Diane Gershon from information provided by the manufacturers. For more information, fill in the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

Mannan binding protein (MBP) is a macromolecular calcium-dependent lectin of approximately 18 subunits, each of 32 kD. Structurally the protein resembles C1q and comprises three distant domains: a cysteine-rich N-terminal domain, a collagen-like domain and a C-terminal carbohydrate recognition domain. The monoclonal antibody to human mannan binding protein (hMBP) that is supplied by Cymbus Bioscience can be used in establishing assay procedures to detect deficiencies in serum levels of MBP (Reader Service No. 118). Such deficiencies may be correlated with opsonic defect and have identified a function of MBP in initiating complement binding to antigenic material rich in mannan or N-acetyl glucosamine. □

ADVERTISEMENTS

Zeinh 1D/2D/DNA/TLC
HAND-HELD SCANNING DENSITOMETER
To quantify:
Gels, blots, TLC, X-Ray film, Photo. 1-D pkg price \$995

8 bits, 60 um resolution, 0-3 OD, sample 11x40 cm. IBM or clones
Options: UV, 2-D, DNA, ELISA, colony counting. \$995 each.
Universal software: for any scanner. Main 1D, 2D, DNA, \$995 each.
Representatives needed, 30% discount of USA price list.

Do-it-all: Departmental, Laser + Video
Micro Densitometry Station
\$ 16,990
Hi-Speed Soft Laser
imaging Video

BIOMED INSTRUMENTS INC. 1020 S Raymond Av. #B, Fullerton
CA 92631. Phn: (800) 927-2290, (714) 870-0290, Fx: 17148706385

Reader Service No.22

PACEMAKERSM

The multiple peptide synthesis and conjugation service

- High purity (up to 95%).
- Quantities for a range of uses.
- Single or multiple peptides.
- HPLC trace with each peptide.
- Terminal modification at no extra charge.
- Conjugation to carrier proteins.

from **£80 per peptide**

AFFINITI Research Products Ltd.
GPT Business Park, Technology Drive,
Nottingham, NG9 2ND, UK.
Tel: +44/0 602 436100; Fax: +44/0 602 436300

Reader Service No.23

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions of the text, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, who are chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive galley proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check their proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few subheadings of two or three words.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the commentary editor in the form of a one-page synopsis. Commentaries are normally between one and four pages of *Nature*.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report.

Most are commissioned but proposals can be made in advance to the News and Views editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words and 5 references.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and three copies are required, each accompanied by artwork. If photographs are included, four sets of originals are required; for line drawings, one set of originals and three good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their manuscript numbers.

Titles say what the paper is about with the minimum of technical terminology and in fewer than 80 characters in the case of Articles and Letters. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs. Numbering of sequences should be in the left-hand margin only, with a single space between rows.

Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent the publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference, including unrefereed abstracts, should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used except for changed addresses.

Acknowledgements are brief; grant and contribution numbers are not allowed.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1137 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT). □